NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

LOOK AGAIN: AN INVESTIGATION OF FALSE POSITIVE DETECTIONS IN COMBAT MODELS

by

Ryan K. Wainwright

June 2008

Thesis Advisor:	Lawrence G. Shattuck
Second Reader:	Christian J. Darken

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** June 2008	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE Look Again: An Investigation of False Positive Detections in Combat Models			**5. FUNDING NUMBERS**
6. AUTHOR(S) Wainwright, Ryan K.			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

This thesis investigates the role of false positive detections in simulated combat environments. Existing combat models tend to overlook or downplay false positive detections. Signal Detection Theory provides the framework for analysis of an observer's hits, misses, correct rejections, and false alarms. The experimenter hypothesized that false alarm rates are a function of observer experience, task instructions, and scene difficulty. In support of this thesis, the researcher developed 24 computer images containing varying numbers of human combatants in an urban environment. Sixteen students at the Naval Postgraduate School volunteered as observers for this experiment. Experimental results revealed that the factors significantly affecting false alarm rates were scene difficulty, task instructions, and the interaction of these two factors. Observer experience was not shown to be statistically significant. Observers given permissive instructions generated up to 3.5 times as many false alarms as did those given restrictive instructions. This experiment showed that the practice of modeling false alarms solely as functions of target and scene characteristics is inadequate. With respect to the generation of false alarms, future combat models must incorporate an assessment of the instructions given to the observer.

14. SUBJECT TERMS False Positive, False Alarm, Signal Detection Theory, Combat Models, ACQUIRE, CASTFOREM, Visual Search, Eye Tracking, Instructions, Combat Experience			**15. NUMBER OF PAGES** 101
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

LOOK AGAIN:
AN INVESTIGATION OF
FALSE POSITIVE DETECTIONS
IN COMBAT MODELS

Ryan K. Wainwright
Captain, United States Army
B.S., United States Military Academy, 1999

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2008

Author: Ryan K. Wainwright

Approved by: Lawrence G. Shattuck
Thesis Advisor

Christian J. Darken
Second Reader

James N. Eagle
Chairman, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This thesis investigates the role of false positive detections in simulated combat environments. Existing combat models tend to overlook or downplay false positive detections. Signal Detection Theory provides the framework for analysis of an observer's hits, misses, correct rejections, and false alarms. The experimenter hypothesized that false alarm rates are a function of observer experience, task instructions, and scene difficulty. In support of this thesis, the researcher developed 24 computer images containing varying numbers of human combatants in an urban environment. Sixteen students at the Naval Postgraduate School volunteered as observers for this experiment. Experimental results revealed that the factors significantly affecting false alarm rates were scene difficulty, task instructions, and the interaction of these two factors. Observer experience was not shown to be statistically significant. Observers given permissive instructions generated up to 3.5 times as many false alarms as did those given restrictive instructions. This experiment showed that the practice of modeling false alarms solely as functions of target and scene characteristics is inadequate. With respect to the generation of false alarms, future combat models must incorporate an assessment of the instructions given to the observer.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

Combat models and simulations play important roles in military training and research. Simulations that include force-on-force maneuvering and engagements contain parameters and algorithms that govern their interactions. Modelers typically take great care to accurately model the physical properties that most affect target detections. Despite their complexity, however, combat models tend to implicitly overlook or downplay "false alarms." Signal Detection Theory succinctly describes the target detection environment replicated in combat models and facilitates analysis of an observer's hits, misses, correct rejections, and false alarms.

Current Army field manuals encourage Soldiers to accept false alarms rather than missed targets. Additionally, most studies in target detection omit the possibility of more than one target per scene. The experimenter hypothesized that false alarm rates will decrease as observer experience increases, as instructions become more restrictive, and as targets become more salient. In support of this thesis, the researcher developed 24 computer images containing varying numbers of human combatants in an urban environment. Sixteen students at the Naval Postgraduate School volunteered as observers for a laboratory experiment.

Experimental results revealed that the statistically significant factor affecting false alarm rates were scene difficulty and task instructions. The interaction between these two factors was also a significant factor. Analysis of hit rates revealed that the only significant factor was scene difficulty. Observer experience was not shown to be significant in either case. Further analysis revealed that when inspecting easy scenes the average false alarm rate of those given permissive instructions was more than 3.5 times that of those given restrictive instructions. For harder scenes, the ratio of the differences is only 1.9.

This experiment showed that the practice of modeling false alarms solely as functions of target and scene characteristics is inadequate. With respect to the generation of false alarms, future combat models must incorporate an assessment of the instructions given to the observer (and possibly training, fatigue, rules of engagement).

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I gratefully acknowledge the assistance of Patrick Jungkuntz and Dr. Robert Easterling in developing the experiment and processing its results.

Above all others on this earth I give my love, thanks, and appreciation to my wife, Mary.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. STATEMENT OF THE PROBLEM

The United States Department of Defense makes great use of combat models. The algorithms of these models vary in complexity and employ algorithmic entities to represent personnel and weapons systems. These models typically contain sophisticated perception algorithms that focus on whether these entities can "see" each other. At a given point in time, a computer model determines visibility of one entity by another by evaluating parameters such as range, azimuth, terrain, and ambient conditions. If visibility exists, then the entity will take action as defined by the algorithm. If visibility does not exist, then the entity will take some other action. Due to the use of fixed values these models often "yield questionable results for individual units" (Correia, 2005). Such algorithms emphasize the physical aspects of target detection while all but neglecting the mindset of the entity. Even the most complex combat models may neglect the possibility of entities mistakenly perceiving other entities that do not exist. These entities therefore do not accurately represent human behavior on the battlefield. The result is that the quality and effectiveness of the combat model are diminished.

In training and in combat, there are many instances when one member of a patrol or crew announces that he "thinks he sees something." For instance, suppose a Soldier in an urban combat situation perceives a person through a doorway. Depending on the tactical situation, he chooses to maneuver, expend ammunition, and communicate with superiors and subordinates. At the very least, he may momentarily stop to consider his options. After developing the situation, the Soldier realizes that the supposed "person" was simply a coat and hat hanging on a hook. He then communicates the discovery to his unit and resumes the mission. While the consequences of this event may appear negligible, they nonetheless influence the mission's ultimate outcome. At the very least, the event caused a minor delay in the patrol's arrival at subsequent locations on the battlefield. Through a combination of training, recent experience, intelligence briefings, adrenaline, and innumerable intangible factors, the Soldier nonetheless "thought he saw something." After further inspection and investigation, however, he determined his initial detection to be incorrect. The Soldier in this scenario experienced what is known as a false positive detection.

Briefly, Signal Detection Theory involves "two possible states of the world, the information, and the decision… The two states of the world are often the absence and the presence of a signal" (Green & Swets, 1988). An observer gathers information through observation and decides to respond to the signal in one way or another. In general, the observer may respond with either "signal" or "no signal" to any of the stimuli. If a signal is present, then the observer's correct declaration of "signal" is deemed a "hit," while the participant's incorrect declaration of "no signal" is deemed a "miss." Naturally, a response of "no signal" when no signal is present is called a "correct rejection." Finally, if the participant's response is "signal" when a signal is, in fact, not present, then this is declared a "false alarm" or "false positive."

Sensitivity is "a measure of the subject's proficiency" to the difference between the two states of "signal" and "no signal," and "is simply the observer's ability to discriminate" between those states (Macmillan & Creelman, 1991). Denoted by *d'*, sensitivity is scaled between zero and one, with an infallible observer having a d' equal to one. Response bias refers to a participant's "willingness to say "yes" [that is, signal present] rather than "no" [that is, no signal present]" (Macmillan & Creelman, 1991), and is also scaled between zero and one. Varying either the observer's sensitivity or bias, or both, changes his resultant rates of hits and false alarms.

The previous example of the Soldier on patrol illustrates the mechanics of Signal Detection Theory. The two 'states' here are whether the person through the doorway is an enemy combatant (a target) or a friendly or noncombatant (a non-target), and whether the Soldier assesses the person as a target or non-target. The resulting four possible combinations of truth/assessment are target/target (a "hit"), target/non-target (a "miss"), non-target/non-target (a "correct rejection"), and non-target/target (a "false alarm"). For a "hit" the Soldier accurately detects an enemy, while for a "correct rejection" the Soldier rightly holds his fire. The "false alarm" of detecting a friendly as an enemy and the "miss" of failing to detect an enemy both carry significant adverse consequences.

Now, suppose the Soldier was told that all enemy combatants are (absurdly) characterized by red shirts and red hats. His resulting sensitivity would be very high, because he is readily able to distinguish enemy combatants from fellow Soldiers and noncombatants. Or suppose that the Soldier was recently reprimanded and demoted after issuing a "false positive" report during a previous patrol. He might favor missing a legitimate target much more than risking another such error. He would require much more certain information about the target before deciding to respond, resulting in a bias against determining that the person in the doorway is an enemy combatant.

By failing to appropriately account for the detections of non-targets, combat models implicitly ignore the possibility of entities generating false positive detections. Attempts to include false alarms in a model's design often consist of inserting additional entities into the model. These entities serve only as a means of generating false alarms and are not otherwise a part of the simulation. While the resulting false alarms may loosely approximate the expected number of false alarms, they are generally not descended from experimental measures of human performance. The goal of combat models is to accurately reflect modern warfare, and better models will result in more success on future battlefields. Recording the responses of Soldiers in target-detection situations similar to those in a modeled environment and designing the model's entities accordingly will improve the resulting combat model.

B. RESEARCH QUESTIONS

The broad question under investigation is, "What is the frequency of false positive perception in unaided visual search for human targets?" Signal Detection Theory provides a broad basis from which to examine this question. Its theoretical structure emphasizes the interactions of bias and sensitivity with respect to target detection performance. The goal of this study, then, is to determine how bias and sensitivity influence false positives when looking for human targets. The following research questions follow: How does real-world training and experience influence an observer's rate of false positive detections? How do task instructions influence a searcher's rate of false positive detections? How does the detectability of a target within a scene affect an observer's rate of false positive detections? These questions touch upon aspects of bias and sensitivity that are likely to be important indicators of an observer's target detection performance.

C. LIMITATIONS OF THE STUDY

The research and experiment are guided toward determining the causes of false positive detections in an infantry combat environment. The observer is a visually unaided human and the target is a human-like figure in a computer-generated combat environment. In particular, the experiment is designed to determine the extent to which an observer's target detection characteristics are the product of bias and sensitivity, and whether combat training or experiences significantly influence the observer's responses. In all cases, the target detection situation is static.

D. THESIS ORGANIZATION

The remainder of this thesis will be organized as follows:

- Chapter II: Literature Review. A look at recent and classic research in the fields of combat models, signal detection theory, military doctrine, and eye tracking.
- Chapter III: Method. Describes the experimental procedures in light of the understanding developed in Chapter II.
- Chapter IV: Results and Analysis. Statistical evaluation of the data.
- Chapter V: Discussion. Qualitative analysis of the data given the statistical evaluation.
- Chapter VI: Conclusion and Recommendations. Discussion of the implications of experimental results on future target detection studies and to current and future combat models.

II. LITERATURE REVIEW

A. CURRENT STATE OF COMBAT MODELS

Combat models can be very powerful in training Soldiers to make informed decisions. For some purposes, high resolution combat models provide the tactical level detail needed to shed light on individual performance and small unit engagements. They allow the representation and tracking of individual entities. For other purposes, the aggregation of hundreds or thousands of personnel and systems in low resolution combat models is more appropriate. These aggregated force simulations often integrate various analytical models. Additionally, the training and reinforcing of tactics and procedures are the intended purposes of some combat models, while others are intended to reveal more about the nature of developing systems or equipment. The nature of the problem being addressed is typically the driving factor behind which type of combat model is more appropriate.

Naturally, accurately representing the capabilities of the various personnel, systems, and units in a combat model is of paramount importance. Otherwise, the simulation is little more than entertainment. Combat modelers therefore take great care in designing algorithms and applying parameters. Depending on the model's intended resolution, some details are represented in total, some are directly represented, and some are ignored altogether. While subsequent verification of combat models tends to be straightforward, developing universal techniques for validation remains elusive. Of great interest in this respect is that programmers of combat models generally find it difficult to model errors.

ACQUIRE and CASTFOREM (Combined Arms Support Task Force Evaluation Model) are two combat models currently in wide use in the United States military. ACQUIRE is a computer model designed to empirically assess target acquisition performance through the use of parameters regarding the target, the sensor, and atmospheric conditions (Army Modeling and Simulation Resource Repository (AMSRR), 2006). While it was originally designed to “predict the probability of detecting a target” with the use of night vision devices, its search and target acquisition methodologies are nonetheless used by several advanced combat models.

The ACQUIRE algorithm uses a single parameter called “n50” to aggregate sensor characteristics such as “observer expectations, training, perception of task, instructions, etc.” N50 is intended to be a “scaling factor representing the point at which 50% of the observer population can perform a target acquisition task” (Mazz, 1998a). Since n50 is a parameter that aggregates a great many sensor qualities and characteristics, some of which are not well suited to quantitative representation, application of n50 to all situations is less than perfectly accurate. Fidelity continues to drop when considering that the entire ACQUIRE algorithm was designed to model observer-sensor characteristics using infrared devices.

For some purposes, the use of n50 in ACQUIRE is wholly adequate. Its gives “predictions that [are], generally, representative of what was measured in the laboratory”(Ratches, Vollmerhausen, & Driggers, 2001). The creators of the algorithms admit, however, that “the selection of the appropriate n50 to a specific situation was not always straightforward,” and that factors such as training, motivation, and reward “have never been incorporated into the model” (Ratches et al., 2001). Additionally, ACQUIRE does not attempt to model the influence of observer criterion or account for false detections. To develop high-resolution combat models that represent unaided detection of human targets, programmers must have access to information which overcomes these current deficiencies. In particular, they must have information gathered from visually unaided human observers that accounts for intangibles such as observer expectations, training, motivation, and reward.

CASTFOREM is used “for weapon systems and tactics evaluation in brigade and below combined arms conflicts.” It provides insight into aspects of combat operations such as force capability, force composition, and system effectiveness (Army Modeling and Simulation Resource Repository (AMSRR), 2005). CASTFOREM employs the ACQUIRE model in its representations of false positive detections. To account for these events, CASTFOREM employs a two-stage strategy. First, it generates a distribution of entities on the battlefield equivalent in form to all others except that these entities are internally labeled as "false targets." Second, it uses the algorithms of ACQUIRE to assess whether the real entity can perceive the false entity.

False personnel targets in early versions of CASTFOREM were "randomly distributed from a Poisson distribution" with a mean density of 0.005 false targets per square kilometer. This factor held only when representing an observer's unaided view of personnel targets. It differed from the representational densities of false personnel targets using heat sensitive vision enhancers (such as Forward Looking Infrared) and from the densities of false vehicle targets in both conditions. Upon placement of a false target into the model, it then used "the same criteria for detection and identification" as for true targets. This allowed for the correct rejection of targets subsequently identified as false (Mazz, 2002). Current publications do not make clear how long an entity will continue to maneuver in response to a false target, or how the entity determines that the false target is, indeed, false.

The largest amount of research into false target densities comes from field experiments with thermal image viewers. An experiment conducted by NVESD gathered data from observers of moving and stationary personnel targets. Another experiment, called Thermal DISSTAF, gathered similar data from observers of vehicle targets. An important conclusion from analysis of these studies is confirmation of their hypothesis that when "expecting a higher density of targets, you should also have a higher density of false targets" (Mazz, 2002).

Research into the false target representation of CASTFOREM revealed a nearly linear relationship between n50, scene clutter, and false targets (Mazz, 1998b). The researchers developed an equation (using values for only n50 and clutter) that generates the mean number of false targets to display in a one kilometer square grid. "These false targets are randomly positioned in the grid and become physical entit[ies] assigned the size and contrast of a tank target. The ACQUIRE model is used to determine if and when an observer detects a false target." The primary research into the applicability of n50 and clutter to determine false target densities was in respect to multiple, individual, fully-exposed ground vehicles. Similar research into the application of these factors in the detection of multiple partially exposed individuals is clearly necessary.

Regardless of the model or method, the resulting parameters influence how often and how long the various entities interact with each other. The models make an attempt to incorporate the possibilities of false alarms and correct rejections. In practice, however, this feature is often disabled by model users because it neither sufficiently represents reality nor aids the user in attaining training objectives. This is indicative of an insufficient or inaccurate application of real-world target detection characteristics to existing combat models.

B. CURRENT DOCTRINE

Current U.S. Army doctrine emphasizes the importance of following a hasty area search with a slower and more deliberate search. Field manuals (designated by ‘FM’, an often hyphenated number, and a formal title) for Soldiers with specialized skills such as snipers, scouts, and special forces, describe this process in nearly the same way. Moreover, the field manual that describes the set of combat skills required by all Soldiers also incorporates this hasty-deliberate process. These manuals reflect the common understanding that some targets are generally more visible than others, and that other targets require much more time and attention before detection.

1. Soldier Skills

FM 21-75: Combat Skills of the Soldier (Headquarters, Department of the Army, 1984) describes basic daytime observation and target detection as a two-step process. The first is a “quick, overall search” of the area of interest “for obvious targets and unnatural colors, outlines, or movements.” The second step is to more thoroughly observe the area of interest in “overlapping, 50-meter-wide strips” from side to side. Upon detecting a possible target, the Soldier is to “search it well.”

To properly search an area in this manner, the Soldier must consider (among other things) position, outlines or shadows, and contrasting colors. Likely places for enemy positions are roads, hilltops, and isolated buildings, in addition to areas containing terrain or vegetation which may provide natural cover and concealment. The outlines of personnel and equipment, including shadows, may also reveal the presence of enemy forces. Additionally, contrasts between background color and the color of uniforms, equipment, and skin may be sufficiently large to enable detection.

Similarly, FM 3-20.8: Scout Gunnery (Headquarters, Department of the Army, August 2005) states that Soldiers trained in reconnaissance techniques and practices begin the target acquisition process with target search and detection. They locate targets using the "ground search techniques: rapid scan, slow (50-meter) scan, [and] detailed search." Rapid scan is an unaided search for obvious signs of enemy, starting in the center of the sector, from nearest to farthest point, and scanning left and right in a similar manner. The "slow scan" refers to the 50-meter overlapping sweeps described earlier, and is usually conducted with the use of vehicle optics or hand-held vision enhancers. A detailed search is more careful and deliberate, and the crew "concentrates on one specific area or location and studies that area intensely."

2. Sniper Skills

The techniques for target detection by snipers are similar to those used by reconnaissance and infantry units. Despite generally having only two Soldiers in a sniper team instead of the nine Soldiers in a typical infantry squad, the techniques are nearly the same. FM 23-10: Sniper Training (Headquarters, Department of the Army, 17Aug94) describes that when occupying a firing position, sniper teams conduct hasty and detailed searches. The hasty search begins with the area nearest the position and extends outward, focusing on "specific points, terrain features, or other areas that could conceal the enemy." The detailed search consists of 50-meter sweeps overlapping by at least 10 meters, also progressing outward. Doctrine calls for three or four repetitions of this hasty-detailed search sequence to reinforce knowledge of the area. Recognizing that prolonged observation greatly reduces effectiveness, and that a team can scarcely afford to be inefficient, the manual recommends that "during daylight, observation should be limited to 10 minutes followed by a 10-minute rest."

Finally, while FM 3-05.222: Special Forces Sniper (Headquarters, Department of the Army, 25April2003) also calls for hasty and detailed searches, it elaborates on the prescribed conduct of both. The hasty search is to last only about 10 seconds. Increasing the likelihood of detection of enemy movement comes by “making quick glances at specific points” rather than sweeping eye movements. Any potential detections are to receive further inspection with binoculars or telescope. Although the detailed search progresses in the same manner as in FM 23-10, the manual highlights several additional factors that enable a sniper team to detect and identify objects in an area of search. Target characteristics that differ from the natural surroundings, such as geometric shapes and contrasting textures and colors, are often the causes of target detection.

3. Implications of Doctrine on False Alarms

Absent from discussion in the manuals is how to determine the identity of potential targets. Knowledge of the intelligence estimate, recent enemy activity, terrain, local culture, and other factors are variables that highly influence the Soldier’s development of a target’s identity. Likewise, since it is impossible to prescribe the most appropriate action upon detection in any and every situation, the manuals naturally omit specific instructions on how to proceed following detection. Regardless, the accepted technique of hasty-deliberate visual search appears to be the primary means of keeping low the number of false positive detections. This appears to stem from the error rates associated with the initial, hasty search. The second search, precisely because it is slower and more deliberate, is therefore a safeguard against proceeding solely on the information gained in the hasty search. False positive detections, then, are presumed to be more likely in the initial search. Additionally, they appear to be not only less likely in subsequent searches, but the information gained in the slower search is assumed to refute or confirm the presence of suspected targets.

The manuals are in agreement in implying that false positive target detections are most likely to occur in an observer’s brief initial viewing of a scene. Observation time is limited by the tactical situation because the enemy is (presumably) simultaneously searching for the observer: detecting the enemy after the enemy detects the observer may very well cost the observer his life. The desires of self-preservation and mission accomplishment are sufficiently strong motivations for observers to remain alert to enemy presence. Accordingly, Soldiers are urged to initially allocate target detection resources primarily toward the methods of the hasty search and to conduct a more thorough, detailed search only if and when the tactical situation permits. In this way, a Soldier conducting a hasty search may quickly find potential targets and decide how to respond. It is only when the Soldier does not find any likely targets that he then proceeds to conduct a deliberate search in hopes of detecting a previously overlooked target. The doctrinal manuals thus appear to reveal an acceptable type of detection error, namely, the favoring of false alarms rather than missed targets. Naturally, a Soldier is more willing to accept a greater number of false positive detections in order to increase the number of hits.

C. SIGNAL DETECTION THEORY

Signal Detection Theory "is a theory about the way in which choices are made" (McNicol, 1972). It encompasses two states and an observer's assessment of those states. While some trials contain only noise, other trials contain a signal in the midst of noise. Observers must make decisions where the full weight of evidence does not fully support a particular hypothesis. An observer's responses to multiple trials may include responses of each of the four types: hits, misses, correct rejections, and false alarms. This information permits calculation of the observer's hit rate and false alarm rate:

$$\text{Hit Rate} = \frac{\text{Number of Hits}}{\text{Number of Signal Trials}} = \frac{\text{Number of Hits}}{\text{Number of Hits + Misses}}$$

$$\text{False Alarm Rate} = \frac{\text{Number of False Alarms}}{\text{Number of Noise Trials}} = \frac{\text{Number of False Alarms}}{\text{Number of False Alarms + Correct Rejections}}$$

Signal Detection Theory commonly uses equal-variance Gaussian curves to represent the distributions of noise trials and signal trials (Figure 1). For the purposes of describing a participant's behavior and predicting future responses, the equal-variance Gaussian detection model is "the most commonly used distribution in signal detection models" (Wickens, 2002). Its two component parameters are sensitivity and bias. Sensitivity, denoted by d', reflects the distance between the means of the two curves and "measures how readily the signal can be detected." A larger value of d' indicates a greater ability to distinguish signal from noise. The observer's criterion, denoted by the

Greek letter, λ, refers to the point at which the observer decides between 'yes, signal' and 'no, noise.' A smaller value indicates a greater willingness to decide 'yes, signal,' regardless of the evidence contained in the trial.

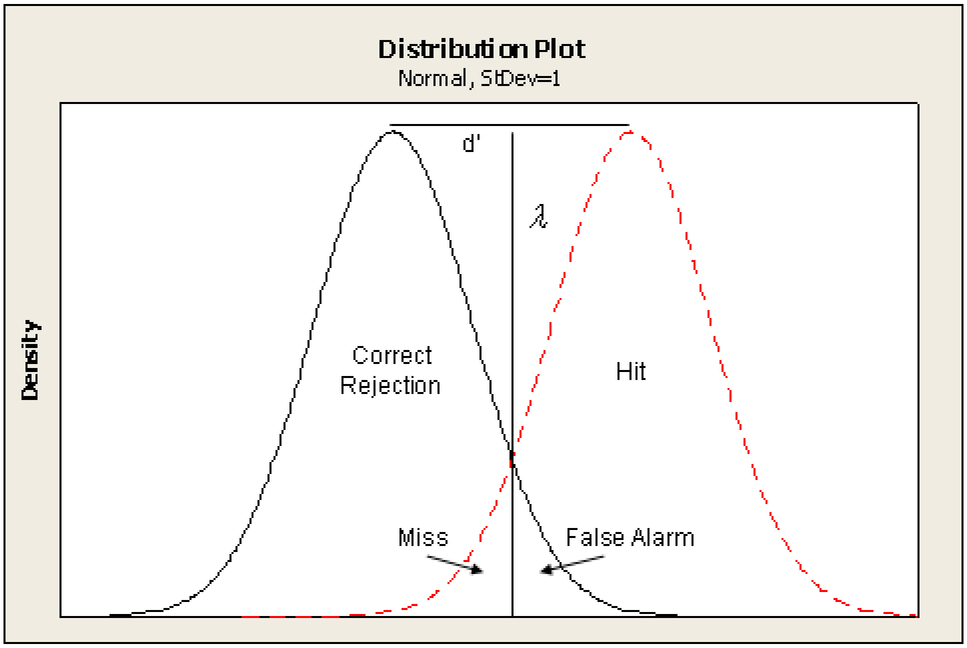

Figure 1. Depiction of the four responses on equal-variance normal distribution curves.

The observer's hit rate and false alarm rate permit calculations of estimates of his sensitivity and criterion. Regardless of the target detection task or the measurements of responses, hit rates and false alarm rates do not immediately provide enough information to thoroughly describe the observer's detection performance. Applying the inverse of the cumulative Gaussian distribution reveals estimates of the parameters:

$$\hat{d}' = z(Hits) - z(FalseAlarms)$$

$$\hat{\lambda} = -z(FalseAlarms)$$

These estimates, along with their associated hit and false alarm rates, more adequately describe the observer's detection and decision processes.

A critical deficiency with $\hat{\lambda}$ is that it includes false alarms while ignoring hits. An observer's criterion crosses the Gaussian curves representing 'noise alone' and 'signal plus noise' each at only one point. β is the ratio of the relative heights of the two distributions at the criterion:

$$\beta = \frac{f_s(\lambda)}{f_n(\lambda)}$$

where

$$f_n(\lambda) = \varphi(\lambda) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(\hat{\lambda})}{2}}$$

is the probability density function of the standard normal distribution at λ (the observer's sensitivity) and

$$f_s(\lambda) = \varphi(\lambda - d') = \frac{1}{\sqrt{2\pi}} e^{\frac{-(\hat{\lambda}-\hat{d}')^2}{2}}$$

is the probability density function of the standard normal distribution at $\lambda - d'$ (the difference between the observer's sensitivity and bias). Scaling β by applying the natural logarithm permits symmetry of the available range of values. An observer's target detection performance is then best described by hit rate, false alarm rate, sensitivity ($\hat{d}'$), and bias ($\ln(\beta)$).

In the context of a visual unaided search for a human target, many factors contribute to the observer's overall sensitivity. The most salient of these factors include environmental conditions and the target's observable shape, color, texture, and orientation. Additionally, the contrast of the target with its immediate background is likely to play a key role in whether an observer detects the target. In this way, the brightness of the target relative to its background becomes a key factor in target detection.

Similarly, an observer's bias reflects a tendency to favor one decision over another. Achieving a goal of maximizing the number of hit targets is a result of the 'cost' the observer assigns to missing a target. Any experimental design that seeks to clearly differentiate between response biases must assign and confer meaningful penalties for either missed targets or false alarms (Green & Swets, 1988). In the urban combat situations represented by combat models, Soldiers rarely have an opportunity to be resupplied with ammunition. As a result, at any point in a combat situation a Soldier must weigh the benefit of firing at a target with the opportunity cost of firing at future targets. In situations where such supply is low, the observer is in a high-bias condition since he knows that he must ‘make every shot count.’ Where there is no such limitation, the observer has fewer reasons to avoid false positive detections and is in a low-bias condition.

In addition to manipulating an observer’s target engagement environment, the measure of an observer’s bias naturally includes other, more complex, cognitive factors. One such factor is that of training and combat experience similar to the target detection tasks in combat modeling scenarios. An observer with extensive training in urban warfare or who has successfully completed many patrols or convoys in a combat environment may apply search patterns and techniques that may be superior to the search techniques of those who do not have such experience or training. Failing to separate the population of observers in this manner has the potential to be a significantly confounding variable.

D. PREVIOUS RESEARCH

Before developing and designing an experimental procedure that involves pertinent aspects of certain combat models, of current military target detection doctrine, and of Signal Detection Theory, a review of previous studies is in order. While investigation into visual tracking began in the 1930s, experiments designed to capture false positive detections began in earnest in the 1960s. In generally, these experiments were conducted as part of wider studies of vigilance and made use of simple stimuli such as letters, lights, and photographs. More recent false positive studies involve specialized virtual reality and point-of-gaze tracking equipment. Gathering information regarding the following areas of experimental research provides great insight into the development of the current experiment: the number and types of participants, the equipment used, the types of stimuli presented, the instructions given to participants, and the experimental procedures.

1. Participants

In terms of participants, laboratory and field eye-tracking experiments focusing on visual target detection tend to exhibit similar characteristics. Of obvious importance is sample size: an experiment with more participants generally results in a data set that is larger and more robust than an experiment with fewer participants. Certainly, a study must gather enough data to achieve statistical significance in order to reach reasonable conclusions. Previous studies that include false positive detections have made use of a range of participants. While the range extends from as few as eight to more than 70, most have data from approximately 20 participants. (See (Reddi & Carpenter, 2000) and (See, Warm, Dember, & Howe, 1997)). Studies that involved few participants tended to build data sets by executing a large number of trials per participant or continuing through multiple days (i.e., repeated measures studies).

Equally important to obtaining a sufficiently large data set is ensuring that the data come from appropriate sources. Previous false positive detection studies generally made great use of students at undergraduate institutions. In these cases, researchers were careful to mention that all participants were volunteers. To ensure that the data were not skewed by sub-normal vision, most researchers mentioned that the visual acuity of each participant was normal or corrected to normal. Some allowed for self-reporting (Martens, 2004), while others made us of standard Snellen charts to make the assessment (Mannan, Ruddock, & Wooding, 1997). Only rarely were the participants compensated for their efforts or was there mention of the experiment having received local ethical committee approval.

Finally, for every study it remains important that the sample of people providing data is truly a subset of the population being modeled. They must represent the type of people who will most benefit from the research. The nature of some false positive detection studies allowed for participants to come from the adult population at large. Other studies, especially those regarding detections of targets in a military field environment, naturally made use of Soldiers. (See (Woodruff, 1986), (Yeh & Wickens, 1998), and (Ozkaptan, 1979).) The choice of participants thus reflected the nature of the study.

One consequence of using military personnel as participants is that it has the potential to influence the results. On one hand, participation may be coerced, leading to half-hearted completion of the task. Since the participants in the aforementioned studies participated in the experiment while performing official military duties, the assumptions of volunteer participation and informed consent may not be thoroughly valid. On the other hand, the experimenter's rank or branch of service may inadvertently influence the participant's perception of the study's importance or applicability. Nonetheless, these studies in no way appeared to be such that Soldiers would have objected.

The choice of participants in the current study reflected the work of previous false-positive studies. All were military students or faculty at the Naval Postgraduate School in Monterey, California. With ages between 27 and 39, these participants had the varied backgrounds and experiences necessary to represent the general military population represented in combat models. Collection of data from participants was conducted with the presumption of professional duty performance. The experiment received the approval of NPS' Institutional Review Board. Participants were volunteers, were naïve to the specific objectives of the study, and had read and signed a statement acknowledging their informed consent (Appendix B) prior to the start of experimental procedures.

2. Equipment

Early false-positive detection experiments made use of pencil-and-paper manual responses. Recent technological developments have "sparked a renewed interest" into the study of eye movements. Due to "more accurate and robust stationary eye trackers" and computer graphics technology "enabling presentation of full color scene images under precisely controlled conditions… the study of gaze control in scenes has recently experienced a rebirth" (Henderson, 2003). It is not surprising, then, to find that the most recent studies involving false-positive detections make great use of eye-tracking systems.

Current eye-tracking systems use the same general principles to measure a participant's gaze. By directing infrared light into the eye, it is possible to measure the distance between its reflection off of the cornea and the center of the pupil. Recording such measurements at known points effectively calibrates the system. The participant is then free to direct his gaze toward the intended scene. Early eye-tracking equipment required immobilization of the participant's head with the use of a chin rest and temple clamps (Mannan et al., 1997). Subsequent versions removed the need for such unnatural restrictions by incorporating hardware and software that measured and accounted for head movement. Researchers have made great use of these eye tracking systems in recent years. (See (Reddi & Carpenter, 2000), (Henderson, 2003), (Martens, 2004), and (Martens & Fox, 2007).) Still, this sort of apparatus required cumbersome and error-prone equipment to be placed on and near the participant.

The most recent commercially available eye-tracking equipment employs the same general technology, but places no hardware on the participant. The system obtains the required information by placing two cameras near the intended scene or image and directing them at the participant's face. Image parallax produces slightly different facial images. By resolving the differences between these images, the software measures and calculates the participant's point of gaze (FaceLab 4, 2006).

Equally important to the accurate eye-tracking of natural human movements is the display of quality images. While early target detection experiments made use of typed text, drawings, or photographs, more recent experiments have used computer-generated images, video, and first-person interactive games. As mentioned earlier, the use of digital imagery can enable the more precise control of potentially independent variables, thereby removing what could otherwise significantly confound the data. The most recent studies have taken full advantage of this capability. These researchers displayed their full-color computer-generated stimuli on high-resolution computer screens or projections. (See (Correia, 2005), (Jones, 2006), and (Martens & Fox, 2007).)

In addition to recording the participant's gaze, there must be some means of recording the participant's response to the stimulus. Just as it is necessary to choose an appropriate set of participants, it is necessary to choose an appropriate method of response. It must not be obtrusive, demand a great amount of learning, or require undue familiarization. Ideally, the participant's response is intuitive, unambiguous, and precise. The masters theses of Correia and Jones describe their respective use of standard modern office equipment: keyboard and mouse, as in (Correia, 2005) and (Jones, 2006). They recognized that the typical American Soldier is very familiar with common computer equipment, and that requiring a simple "click" or press of a button is almost a second-nature task.

To assist in building upon the body of eye-tracking knowledge, the choice of equipment for this eye-tracking experiment must incorporate the knowledge gained from previous studies. By recording eye movement with unobtrusive equipment, presenting carefully crafted and high-resolution color images presented on high-quality display screens, and recording participants' responses in a way that does not unduly influence eye movements will provide the most representative target detection data possible.

3. Stimuli

Previous experiments employed a wide variety of stimuli in attempts to learn about and model human behavior. In most cases, accuracy (hit rate) and time until required response were the dependent variables. While these provided the information necessary for their particular purposes, such studies often failed to record the precise type of error (that is, whether an error was a miss or a false alarm). The method of data collection thereby lost information that would otherwise shed light on the subject of false positive detections. Nonetheless, methods employed in previous studies are instructive in the design of the current study.

Whether making use of photographs or computer-generated imagery, a common theme among stimuli used in target detection studies is the presentation of "natural-looking" scenes. Capturing target detection and eye movement characteristics of people while looking at abstract or nonsensical images would hardly be expected to help improve combat models. Thus, developing imagery that will help elicit appropriate responses from participants is of prime importance.

Experimenters used stimuli that were appropriate to the aims of their particular studies. While investigating the target detecting characteristics of Army helicopter pilots, Ozkaptan presented still images representing in-flight detection of "a single 20-foot military tank in various field locations" (Ozkaptan, 1979). Yeh displayed various military targets in hilly terrain, having artificially adjusted the intensity of the targets to maintain similar contrast ratios among scenes (Yeh & Wickens, 1998). More generally, Mannan sought greater understanding of eye fixations by presenting multiple images of "natural landscapes, buildings, vehicles, and human figures and faces" while varying the level of image quality (Mannan et al., 1997). Lastly, research by Henderson emphasizes not only the natural development of an artificially developed scene, but also the natural placement of the target. He notes that both factors are important because even the first fixation "provide[s] important information about where a particular object is likely to be found" (Henderson, 2003). From these observations it is clear that the stimulus set must be carefully chosen and developed.

The most recent target-detection research conducted at NPS is discussed in the masters theses by Correia and Jones. Both used imagery based on the Delta 3D open-source simulation engine and developed by the MOVES Institute. The scenes are still images containing a single human-like figure in urban and rural environments. (A few scenes did not contain a target.) They were able to vary several important scene and target characteristics, such as figure position and orientation, level of obscuration (smoke or fog), and time of day. (See (Correia, 2005) and (Jones, 2006).)

4. Instructions

As with all experimental situations involving human participants, the instructions given to participants can have significant effects on performance. Consequently, the resulting body of data may contain a large amount of what would be otherwise avoidable variance. Most researchers attempt to remove this source of variability by delivering instructions to each participant in the same manner, such as through written or videotaped instructions.

The complexity of instructions in an experiment naturally depends on many factors, especially the objectives of the experiment, the stimuli, and the means of recording the data regarding the dependent variables. For some experiments in target detection, the instructions were very simple. For example, the experiment by Martens instructed participants to simply behave "as in real life" (Martens & Fox, 2007). Since the experiment involved a vehicle driving task that replicated daily activities, this sort of broad approach was appropriate. Other experiments included instructions that were quite broad. In a study of fixations on still images, Mannan instructed observers to "examine the images as carefully as possible in order to answer unspecified questions about them" (Mannan et al., 1997). And while the goals of experiments by Correia and Jones were to investigate detections of human-like targets, the language they used in their instructions to participants was vague in that they only implied the presence or absence of a target in each scene. (See (Correia, 2005) and (Jones, 2006).)

Providing guidance to participants in a target detection task is critical because all observers bring their own biases to the target detection situation. Left without influence, the observer simply decides whether to emphasize either speed or accuracy. The Woodruff study, which sought to relate search performance and observer expectations, urged rapid target detection and instructed participants to respond when "50% sure there is a target where you are looking... then immediately continue searching for other targets" (Woodruff, 1986). Alternatively, other studies placed greater importance on responding correctly. When the task is to respond as accurately as possible, the times from the presentation of the target until the participant's response are often not even recorded (Martens, 2004). In an effort to discourage participants from guessing at target locations, Jones went so far as to explicitly inform them that the "experiment does not take time into consideration" (Jones, 2006).

While encouraging either accurate or speedy responses is acceptable in experimental situations, such restrictions are unrealistic in most real-world situations. Whether inspecting products for deficiencies on an assembly line or searching for enemy combatants in a high-intensity combat zone, accuracy and speed both are important attributes. As discussed earlier, Signal Detection Theory reveals that simultaneously achieving both of these objectives is highly improbable. Nonetheless, striving for both objectives is, in most situations, both encouraged and expected. While researching the influence of urgency on decision time, Reddi and Carpenter designed an experiment with two such conditions. They instructed participants to either "move at a comfortable pace and be as accurate in responses as possible" or "respond as rapidly as possible, worrying less about making mistakes." At the end of the study, the researchers concluded that "instructing subjects to respond either more carefully or more hastily is equivalent to altering the threshold…at which a response is initiated" (Reddi & Carpenter, 2000).

Ozkaptan conducted an experiment with similar instructions, with an additional 'neutral' condition. Each instructional set contained verbiage that enabled the participant to more readily accept the instructional bias. He concluded that "instructional set is an important determinant of performance during target acquisition with respect to reaction time and the number of hits" (Ozkaptan, 1979). Even very similar studies employ these opposing strategies. While Correia instructed participants to find the target "as quickly as possible, but take all the time needed" (Correia, 2005), Jones instructed participants to "be as accurate as you can, but be sure you see or don't see the character before taking action" (Jones, 2006). At the very least, instructing participants to respond either accurately or quickly is equivalent to manipulating an independent variable.

At the most fundamental level, instructions are given to influence responses in a particular way. Experimenters develop instructions with the intention of conforming responses to within intended bounds, making subsequent data analysis tractable. As a result, the participant develops expectations that affect subsequent responses to stimuli. In a study involving theoretically trained and untrained observers, Nieman et al found that accuracy of judgment decreases when theoretical expectations are violated (Nieman, Roberts, & Kantner, 1983). The participant's expectancy was a result of formal instruction prior to presentation of stimuli. Similarly, the experiment conducted by Yeh demonstrated that expectancies improve search performance when a target is present, but inhibit search performance when no target is present (Yeh & Wickens, 1998). Here, expectancy was the result of visual cueing throughout the session. Pavel recognized that "change in sensitivity depend on the attentional instructions as well as on the complexity of the tasks" (Pavel, 1987). Regardless of the manner of developing expectancy, the mindset of the participant as manipulated by the experimenter clearly influenced results.

Various studies have shown how expectation yields levels of responses. While the previously mentioned studies describe generally discrete response differences, other studies reveal that expectancy can influence response in a more continuous manner. In one experiment, See et al gave participants of four groups varying levels of "probability of signal occurrence they would encounter during the session." She found that "bias was most conservative at a probability of .05 and became progressively more lenient as the probability of signal occurrence increased to .75" (See et al., 1997). Thus, a participant is more likely to respond to a stimulus when he or she already expects it to be there. While See's study dealt with the presence and absence of targets, Pavel demonstrated that the expected location of a target influences its detection. He concluded that "expecting a stimulus at a particular location enhanced perceptual sensitivity for that location," and that such sensitivity decreased as the target's distance from the expected location increased (Pavel, 1987).

As discussed earlier, one of the most intuitive means of influencing target detection bias in a combat situation is to vary the implied 'supply' of accurate target detection responses. An observer who seeks to maximize the number of hit targets will respond to stimuli in a way that appropriately reflects such a limitation.

5. Procedures

The experimental procedures used in target detection and eye tracking research vary widely. As with the other aspects previously discussed, the chosen procedures reflected the intent of the overall study. The choice of equipment generally appeared to have the greatest influence in this regard.

The studies by Correia and Jones are those which are most similar to the current effort. To achieve consistency of data, the experimenters read scripted instructions before each participant's session. They also conducted a brief practice session that enabled the participant to become familiar with the laboratory environment, the general format of the stimulus, and the equipment. Having prepared each participant in the same manner, they then presented the stimuli and recorded results. While most scenes contained a single target, some scenes intentionally did not. The sequence of images was identical for all participants, and all participants were permitted to proceed at their own pace. Following completion of the formal data collection, the experimenters asked participants to complete a brief questionnaire.

With regard to image sequence and exposure time, earlier experimenters used different approaches. The images in the Mannan experiment were briefly displayed for only three seconds and in a random order. Additionally, the experimenters presented a central fixation target "which constrained eye movements to start at the center of the screen" (Mannan et al., 1997). They then presented the next image 60 ms after removing the central fixation target. While Ozkaptan also presented scenes with and without targets, he set the exposure time at 30 seconds (Ozkaptan, 1979). Varying exposure time is one way to influence the participant's reaction to each stimulus. Some researchers do so as a means of manipulating an independent variable. Random presentation of images serves to counterbalance the effects of learning that may take place as the experimental session progresses.

E. IMPLICATIONS AND HYPOTHESES

Any meaningful attempt at investigating target detections must adapt previous studies to the particular population of interest. Since current doctrinal approaches to target detections imply preference of false alarms over missed targets, the methods of representing false alarms in combat models ought to be more thorough and robust. As such, the following laboratory experiment builds upon previous target detection studies by investigating whether false alarm rates decrease as observer experience increases, as task instructions become less restrictive, and as target salience increases. Based on the literature discussed above, the following hypotheses were proposed: Hypothesis #1) The rate of false positive detections decreases among those with high levels of target detection training and experience; Hypothesis #2) The rate of false positive detections increases when detection responses are less restrained; Hypothesis #3) The rate of false positive detections decreases when searchers can more easily detect targets.

Whether the goal of a particular combat model is to train and develop its users or to shed light on the complex nature of modern warfare, the model must accurately reflect the detections of targets. As discussed previously, combat models often implicitly ignore the possibility of entities generating correct rejections and false positives. The next chapter details the researcher's experimental methodology. By recording participants' hits, misses, correct rejections, and false alarms in target detection situations similar to those in a simulated environment, programmers will be better able to improve tomorrow's combat models.

THIS PAGE INTENTIONALLY LEFT BLANK

III. METHOD

A. PARTICIPANTS

In this study, 16 students at the U.S. Naval Postgraduate School were the ‘experimental units’ and were evenly ‘blocked’ according to levels of experience. ‘Experience’ in this study referred to military target detection training, advanced infantry training, and experience in ground combat or convoy operations. Voluntary completion of a questionnaire (Appendix A) screened these military officers based on experience. Gathering these responses and subjectively arranging them from most experienced to least experienced revealed a wide breadth of experience among participants. Those candidates without infantry training or experiences were considered to be inexperienced. Those having completed over 100 urban combat patrols or convoys were considered to be experienced. Those who had only some infantry training or experience and could not be easily classified were not selected for further participation.

The sixteen participants represented all branches of the United States Department of Defense (10 Army, 1 Navy, 3 Marine Corps, 2 Air Force). All but one was male, and exactly half had received surgery to correct visual deficiencies. The average age of participants was 34.1 years (standard deviation 3.5), and the average years of commissioned service was 11.7 (standard deviation 3.7). All participants reported playing first-person shooter video games less often than once per month (the lowest response level).

Prior to data collection, the experimenter ensured that each participant had natural or corrected visual acuity of at least 20/40 and was not color blind. Covering one eye, participants read aloud the smallest line possible from a standard Snellen eye-exam chart on a wall twenty feet away. Participants who could accurately read text corresponding to at least 20/40 vision with both eyes independently were considered able to proceed. Admittedly, this approach is far from truly diagnosing a participant’s precise visual acuity. Nonetheless, the experimenter gained confidence that the observer could effectively identify small features on the stimuli. Figure 2 is a replica of the Snellen chart used to assess each participant's visual acuity.

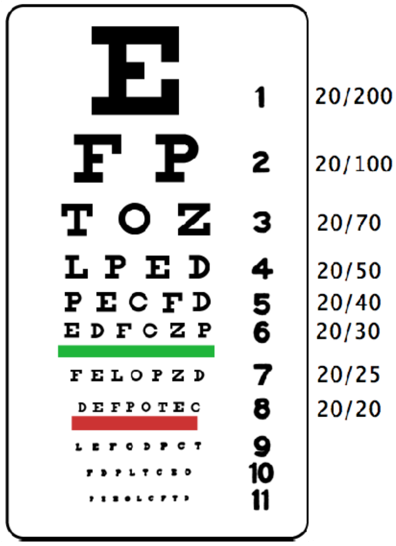

Figure 2. Sample Snellen chart. [From Wikipedia, February 2008]

Normal discrimination of colors is important to effective target detection. Although various means exist to detect and classify color blindness, the Ishihara tests are recognized as an expedient way to broadly assess deficiencies. The experimenters presented the color discrimination slide in the "Optec 2000 Vision Tester – Industrial Model" and had participants read aloud the number clearly contained in six separate Ishihara images. For the purposes of this experiment, accurate identification of the number in the first five images (the sixth did not contain a number visible to those with normal color vision) provided sufficient reason to declare the observer not color blind. Figure 3 is similar to the images used in the test for color-blindness.

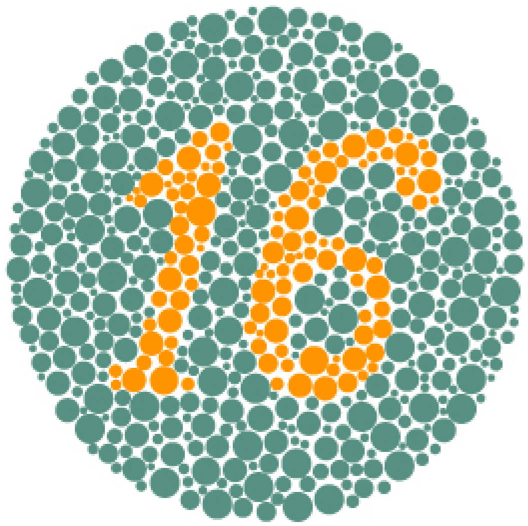

Figure 3. Sample from an Ishihara test. [From Wikipedia, December 2007]

B. EQUIPMENT

The eye-tracking equipment used in the experiment was FaceLab 4, a product of Australia-based Seeing Machines, Ltd. As a “passive measuring device” that employs table-mounted cameras to capture and analyze the characteristics of a participant’s face, FaceLab is nonintrusive and “enables the user to work as they usually would” (FaceLab 4, 2006). This system allows participants to move away from and return to the experimental area without requiring recalibration. Other eye-tracking systems require the participant to wear bulky headgear which holds various lenses, cameras, and wires. Such apparatus generally interferes with natural movement and encourages the experimenter to design the trials such that rest breaks are not included. This system permitted recording of the participant’s gaze on a computer monitor with an accuracy of approximately one degree error.

The stimuli were presented on a 24” (measured diagonally) Dell flat screen computer monitor with 1920x1200 resolution. The window-less experiment room was dark except for the indirect glow created by the laptop computer administering the FaceLab equipment. This laptop was positioned such that it did not produce glare or excessive reflection onto the participants.

In addition to eye movements, participants responded to the stimuli with hand movements. Participants kept one hand on a standard two-button computer mouse and the other on a standard computer keyboard. With the mouse they moved a white arrowhead-shaped cursor on the screen to the location of a suspected target and left-clicked upon it. Participants did this for all targets detected in each scene. Upon completion of a scene's search the participant proceeded to the subsequent scene by pressing the Enter key.

C. STIMULI

The stimulus package used in this target detection research was the result of a collaborative effort. Mr. Michael Dunhour developed the basic structure of the scene-generating software with the use of Delta-3D, an open-source computer game development engine. Mr. Daniel McCue developed the software that enabled the placement of targets, the adjustment of scene conditions, and the recording of mouse-click data. Since previous studies tend to highlight the importance of having a large number of target detection opportunities, the experimenter used these products to develop 24 unique scenes.

Having multiple targets in each scene was an important factor in this analysis. Previous target detection studies rely primarily on "yes-no" or "force choice" tasks. This study took advantage of the most recent eye-tracking technology to obtain detailed information about the participant's behavior while looking for an unknown number of targets in multiple scenes.

The experimenter designed all scenes with at least one target per scene. Presenting even one scene without any targets would have the potential to influence future behavior away from responding with mouse clicks. Regardless of the participant's bias, it would effectively make him require a greater amount of information before deciding on the presence of any target. This would cause the participant to be less willing to respond to ambiguous targets in subsequent scenes. Since the goal of this research was to examine errors in target detection, it was important to encourage, rather than discourage, target detection events.

The experimenter also designed the stimulus package so as not to inadvertently bias the participant. While there is no upper bound on the number of possible targets per scene, the creation of each scene must nonetheless reflected a balance between containing many targets and enabling accurate capture of visual and manual responses. The gaze-tracking limitations of the eye tracking system thus limit the maximum number of targets per scene. As a result, the practical limit to the number of targets in scenes was six.

In the context of searching for multiple targets in a combat situation, the observer's sensitivity is the product of multiple factors. Such factors include the amount of the target within view of the observer, which part of the target is observable, the brightness of the target itself, and the contrast of the target to objects in the foreground and background. The experimenter developed and conducted a small-scale pilot study designed to determine the best ways to manipulate the target saliency, to develop appropriate biasing instructions, and to refine experimental procedures.

The pilot study consisted of four participants viewing 27 unique scenes, each containing between one and six targets. Two of the participants received instructions intended to result in high biases while the others received instructions intended to convey low biases. Participants searched for targets and signaled detection by moving a mouse cursor to the suspected target and clicking once. Analysis of target detection results at the conclusion of the pilot study revealed scene and target characteristics obviously that obviously correlated with trends in hits, misses, and false alarms.

The experimenter then created and developed a set of scenes that balanced not only the total number of easy and hard scenes but also the total number of targets contained within them. While displaying 24 scenes each with six targets would produce a many target detection opportunities, it would also influence the participant's target detection strategy. Presenting scenes containing up to six targets while favoring one format (that is, having most scenes contain one or five or six targets) would likewise inappropriately bias the participant. Randomizing the presentation of an equal number of scenes containing a given number of targets avoided this potentially confounding influence. For this reason, the stimulus package of 24 scenes consisted of one target in each of two easy and two hard scenes, two targets in each of two easy and two hard scenes, and so on up to six targets in each of two easy and two hard scenes. Twelve easy scenes thus contained a total of 42 targets, as did twelve hard scenes. By the conclusion of the stimulus package each participant had been presented 84 targets.

D. INSTRUCTIONS

Having categorized the experience level of participants, the next step was to randomly assign them to groupings of high and low biases. Describing the experimental session in terms of a tactical situation provided the participant with an intended mindset. The precise instructions given to participants thus depended on the participant's experimental group. One group received an instruction set designed to elicit responses arising from a high bias condition, while the other group received an instruction set intended to influence participants toward a lower bias. Participants remained unaware of the different instructions, but all participants received the same information about the general purpose of the study, the equipment to be used, and the general format of the stimuli. Copies of the written instructions given to each participant are contained in Appendices D and E.

A high bias condition is one in which the participant adopts a characteristically 'conservative' approach to responses, while a low bias condition is generally more 'risky' in responses. In order to accept fewer false alarms an observer with a high bias effectively moves his criterion further to the right on the plot of the two Gaussian distributions (but also fewer hits). Given an observer's sensitivity, increasing the hit rate simultaneously increases the false alarm rate. An observer with a low bias behaves in an opposite manner: he moves his criterion further to the left, thereby allowing more hits but also accepting more false alarms.

Participants assigned to the low bias condition were given very liberal (afterwards labeled, "permissive") firing instructions. They were told that only enemy combatants were within the scenes and that the current tactical situation and rules of engagement did not require them to be certain of a target's identity before firing. They were specifically not limited in the number of 'shots' allowed per scene, but were instructed to not take ‘wild shots.’ There was no other mention of penalties for either misses or false alarms. Low-bias participants were therefore expected to have relatively high hit rates and high false alarm rates.

Participants assigned to the high bias condition were given more conservative (afterwards labeled, “restrictive”) firing instructions. Their instructions allotted only a limited, but varying and unknown, number of shots per scene. To ensure that participants had a rather high confidence of a target’s identity, they were told that the number of recordable shots in a scene equaled the number of targets actually in that scene. They were also notified that any subsequent shots would not be recorded. (Naturally, experimenters recorded all shots, not just the ‘recordable’ shots.) Keeping participants unaware of the true number of targets in any given scene discouraged guessing and encouraged them to require a high degree of confidence before responding with a mouse click. High-bias participants were therefore expected to have lower hit rates and lower false alarm rates.

For each combination of experience level and type of instruction, the participant was presented a stimulus package that contained targets both easy and hard to detect. Figure 4 presents the various combinations of experience, instruction, and ease of detection.

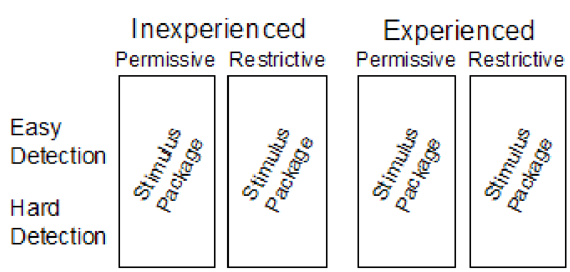

Figure 4. Experimental structure.

E. PROCEDURES

The experiment began with orienting the participants to the laboratory and obtaining their informed consent to take part in the experiment (Appendix B). Participants completed a brief pre-trial questionnaire (Appendix C) and simple tests for visual acuity and color blindness. Participants read the assigned instructional set (Appendices D and E) and verbally answered scripted questions about specific details contained in the instructions. Experimenters then calibrated the FaceLab system to the participant and conducted a familiarization and practice session. This practice session consisted of three short target detection scenarios (three scenes each) immediately followed by a report of the participant's performance. In this manner, participants were further influenced toward accepting the intended biasing condition. Experimenters asked scripted questions appropriate to the given biasing instructions and specifically confirmed the participant's readiness before proceeding to the main trial sequence.

An equal number of participants were classified as having either high experience or low experience. Within these experience groups an equal number of participants were then assigned to either a high bias or low bias condition and given corresponding instructions. Although each participant was easily able to accurately describe his own level of experience, he was ignorant of the intentional differences between the levels of target detection difficulty and the differing instructional conditions.

Researchers presented the same 24 target detection scenes to all participants. The stimulus package included an equal number of easy and hard scenes. These scenes were randomly distributed throughout the stimulus package for each participant. After the training sessions and immediately prior to data collection, experimenters reminded the participant of the unchanging natures of instructions and procedures. Participants then completed, without interruption, the target detection task in the 24 scenes of the stimulus package. Participants were allowed to search each scene for up to 15 seconds, but were permitted to proceed to the subsequent scene at any time. Afterwards, all participants appeared to readily understand the importance of not compromising future data collection efforts by discussing the experimental session with others.

IV. RESULTS AND ANALYSIS

The data recorded from each participant during the stimulus presentation included the number of hit targets, the number of missed targets, and the number of false positive detections. The FaceLab eye-tracking system also recorded the location of the participant's gaze throughout the data-collection session. The system captured and recorded the x- and y-positions of the observer's gaze on the stimulus display screen at a rate of 60 hertz.

A participant's hit rate is dependent on the number of hits and the number of total targets presented. As the designer of the stimulus package, the experimenter knew precisely how many targets it contained. Determining whether a participant's response was a hit, miss, or false alarm was then a relatively simple manner. As a result, the hit rate was easy to obtain.

The determination of a participant's false alarm rate, however, was slightly more complicated. As written in Chapter II, the false alarm rate is the number of false alarms divided by the number of noise trials. That is, it is the quotient of the number of false alarms divided by the total number of false alarms plus correct rejections. The participant's actions with the mouse do not capture the number of correct rejections, and since all scenes had at least one target, the number of 'noise alone' scenes was zero. In situations of multiple targets per scene, the number of noise trials becomes dependent on the participant's gaze. Correct rejections are thus fixations on non-targets followed by a decision to not click upon the object of the fixation.

The data obtained via the FaceLab system permitted calculation of a participant's fixations in each scene. There is no widely accepted definition of either fixations or saccades in terms of distance, angles, or other objective parameters. For the purposes of this study, saccades were presumed to have begun when the observer's point of gaze moved more than the length of four pixels between consecutive recordings. Having defined saccades, fixations naturally followed. Overlaying the set of resulting fixations (numbered sequentially) onto its corresponding stimulus scene enabled the distillation of these calculated fixations into correct rejections according to a set of ‘rules.’ In order for a fixation to be counted as a correct rejection, it must have met all of the following conditions:

- A correct rejection is a fixation farther than approximately one degree visual angle from any target
- A correct rejection is a fixation farther than one degree visual angle from the location of any mouse click
- A single correct rejection is recorded when multiple fixations appear in sequence within approximately one degree visual angle

Given the maximum 15 seconds permitted to search each scene, these analytical conditions result in reasonable numbers of fixations and correct rejections for each scene and for each participant.

A. RAW DATA

Experimental design and data collection permitted calculation of each participant’s hit rate and false alarm rate. From these two terms come individual measures of each participant’s measures of sensitivity and criterion. All such values were separated by performance on ‘easy scenes’ and ‘hard scenes.’ Tables 1 and 2 display each participant’s individual hit rates and false alarm rates, respectively. Figures 5 and 6 graphically display the same information.

Hit Rate		Low Experience				High Experience			
			Permissive		Restrictive		Permissive		Restrictive
Easy Scenes	Participant Number	32	0.976	31	0.810	21	0.952	22	0.905
		37	0.976	36	0.952	33	0.929	34	0.905
		42	0.905	43	0.929	35	0.952	44	0.929
		46	0.857	45	0.952	52	0.952	51	0.952
	Mean		0.929		0.911		0.946		0.923
	StDev		0.058		0.068		0.012		0.023
Hard Scenes	Participant Number	32	0.714	31	0.429	21	0.714	22	0.619
		37	0.714	36	0.738	33	0.524	34	0.595
		42	0.476	43	0.619	35	0.571	44	0.714
		46	0.643	45	0.500	52	0.643	51	0.690
	Mean		0.637		0.571		0.613		0.655
	StDev		0.112		0.136		0.083		0.057

Table 1. Hit rates for each participant in easy and hard scenes.

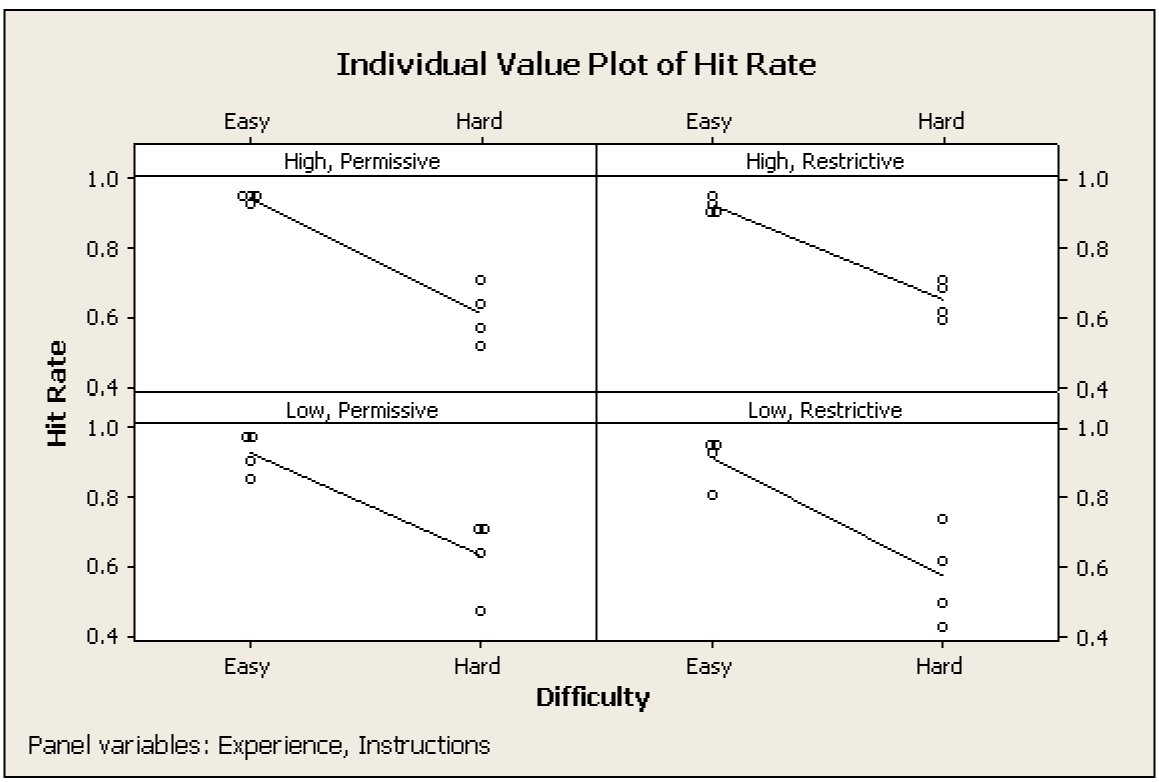

Figure 5. Plots of hit rates on easy and hard scenes

False Alarm Rate			Low Experience				High Experience		
			Permissive		Restrictive		Permissive		Restrictive
Easy Scenes	Participant Number	32	0.056	31	0.024	21	0.092	22	0.008
		37	0.045	36	0.016	33	0.058	34	0.010
		42	0.029	43	0.017	35	0.012	44	0.022
		46	0.096	45	0.022	52	0.060	51	0.010
		Mean	0.057		0.019		0.056		0.013
		StDev	0.029		0.004		0.033		0.006
Hard Scenes	Participant Number	32	0.035	31	0.059	21	0.094	22	0.037
		37	0.078	36	0.047	33	0.092	34	0.026
		42	0.022	43	0.022	35	0.035	44	0.038
		46	0.145	45	0.037	52	0.070	51	0.030
		Mean	0.070		0.041		0.073		0.033
		StDev	0.055		0.016		0.027		0.006

Table 2. False alarm rates for each participant in easy and hard scenes.

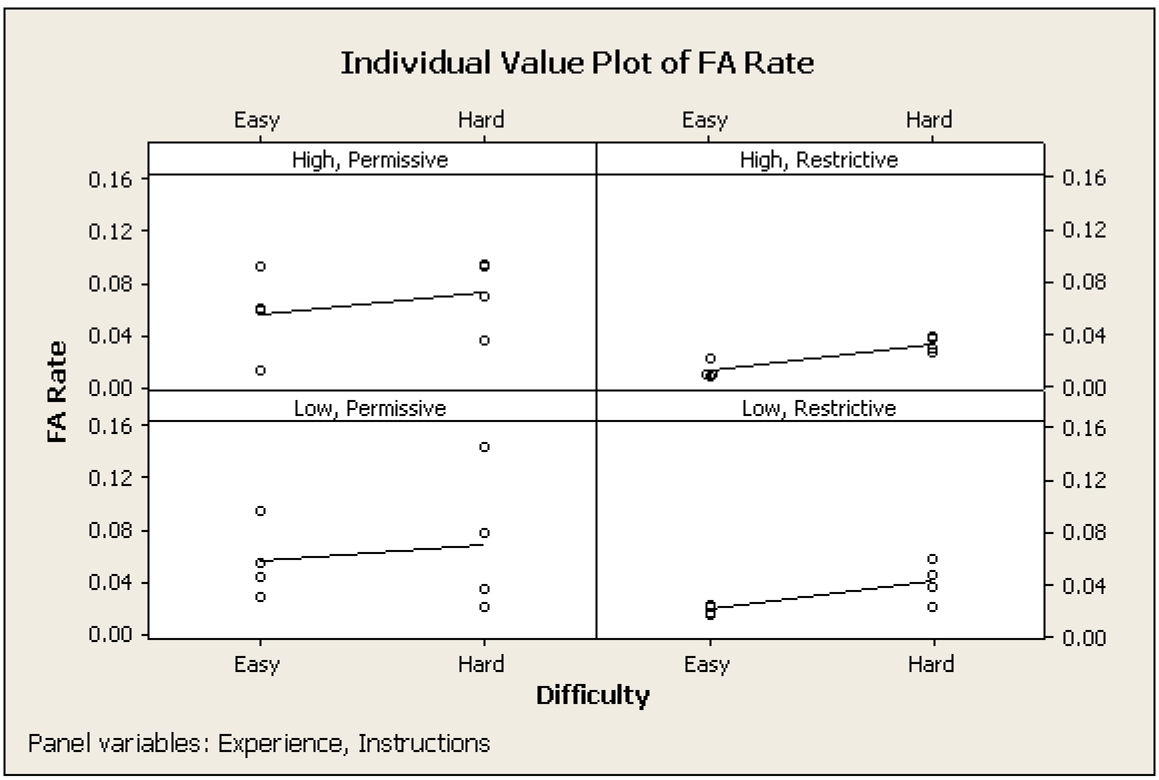

Figure 6. Plot of false alarm rates on easy and hard scenes.

Initial analysis of the hit rates of the 16 participants shows that the mean hit rates with respect to easy scenes falls between 0.91 and 0.95, depending on the group's experience and set of instructions. This seemingly small range of means betrays a relatively wide range of standard deviations: 0.01 to 0.07. Conversely, the ranges of means and standard deviations for hit rates in hard scenes are larger and somewhat wider (mean range: 0.57 to 0.66, standard deviation range: 0.06 to 0.14). Regardless of the differences between easy and hard scenes, the overall variability of standard deviations effectively shows that an assumption of equal variances in this raw data is untenable. Review of the false alarm rates showed relatively wide ranges of means in both easy scenes (0.01 to 0.06) and hard scenes (0.03 to 0.07). Similarly with respect to false alarm rates, the ranges of standard deviations were quite broad in both easy scenes (0.004 to 0.033) and hard scenes (0.006 to 0.055). Analyses of variance (ANOVAs) revealed whether observer experience, scene difficulty, and task instructions significantly influenced the hit rates, false alarm rates, sensitivity, and bias. All subsequent plots and ANOVA tables came from Minitab® 15.1.1.0, 2007.

B. ANALYSIS OF HIT RATES

As noted above, the variability of the raw data cast doubt upon assumptions of normality and inhibited the use of ANOVA. Transforming all hit rates by the logit function permitted further analysis by strengthening the assumption of normality and making variances approximately equal. The transformation brought central data points closer to the center line and brought the errant tail value into the 95% confidence interval. Figure 7 shows the normality probability plots of hit rates before and after the transformation.

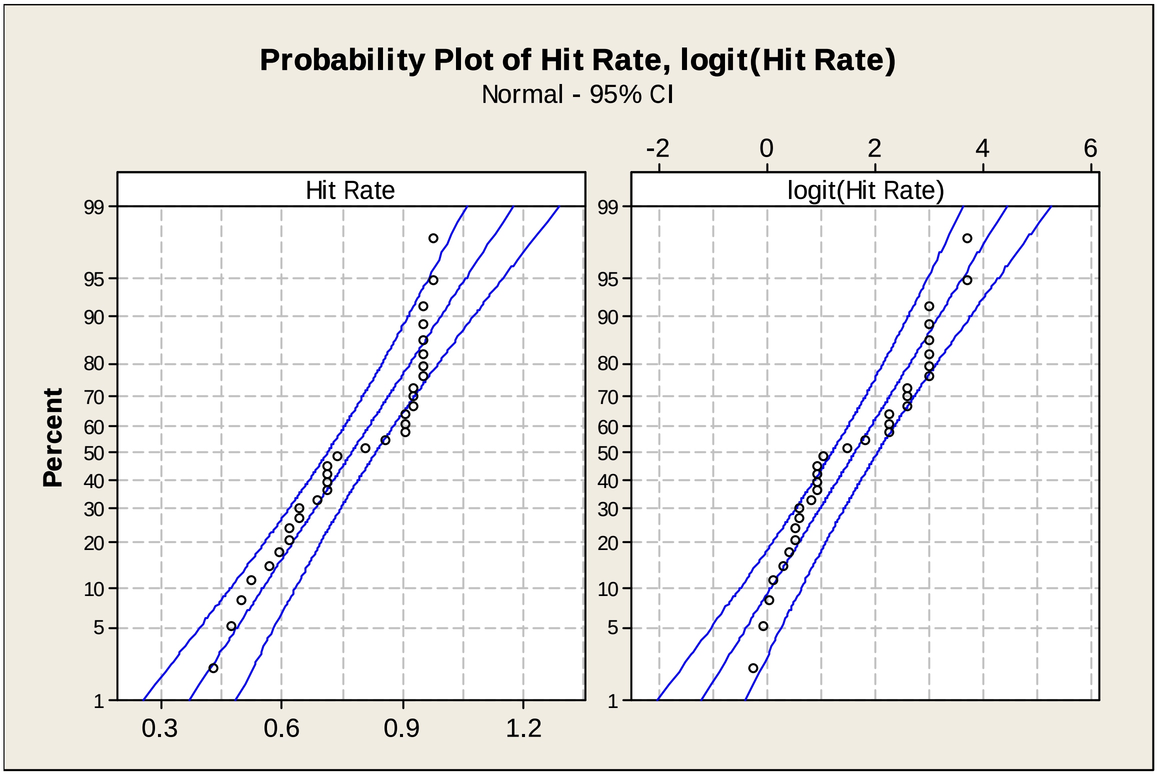

Figure 7. Normal probability plots of Hit Rate and logit(Hit Rate).

With respect to hit rates, the ANOVA table (Table 3) revealed that the only significant factor was Difficulty (p-value <0.001). There were no apparent significant two-way or three-way interactions. As expected (and intended), Difficulty was the only significant factor in determining hit rates. The initial portion of Table 3 lists each factor's type, levels, and values. This information remained unchanged and will be omitted in subsequent ANOVA tables.

General Linear Model: logit(Hit Rate)

Factor	Type	Levels	Values
Experience	fixed	2	High, Low
Instructions	fixed	2	Permissive, Restrictive
Difficulty	fixed	2	Easy, Hard
Participant(Experience Instructions)	random	16	21, 22, 31, 32, 33, 34, 35, 36, 37, 42, 43, 44, 45, 46, 51, 52

Analysis of Variance for logit(Hit Rate)

Source	DF	Seq SS	Seq MS	F	P
Experience	1	0.0352	0.0352	0.07	0.792
Instructions	1	0.3431	0.3431	0.71	0.416
Experience*Instructions	1	0.0990	0.0990	0.20	0.659
Participant(Experience Instructions)	12	5.8022	0.4835	3.92	0.013
Difficulty	1	38.3353	38.3353	311.14	**0.000**
Experience*Difficulty	1	0.0183	0.0183	0.15	0.707
Instructions*Difficulty	1	0.2014	0.2014	1.63	0.225
Experience*Instructions*Difficulty	1	0.1006	0.1006	0.82	0.384
Error	12	1.4785	0.1232		
Total	31	46.4135			

Table 3. ANOVA of logit(Hit Rate)

A plot of the main effects (Figure 8) shows how scene Difficulty contributed most to an observer’s hit rate while Experience and Instructions contributed very little.

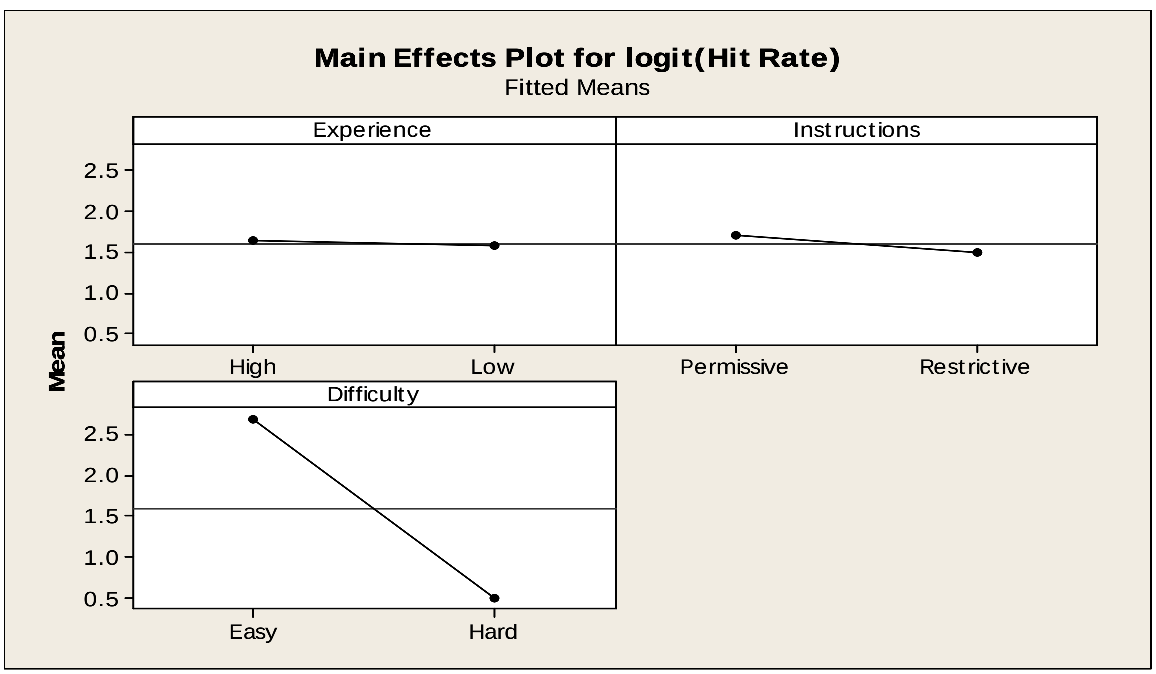

Figure 8. Main effects plot for logit(Hit Rate).

Similarly, a plot of the factor interactions (Figure 9) shows that Difficulty contributed most to hit rates and that there was no evidence of any significant factor interactions.

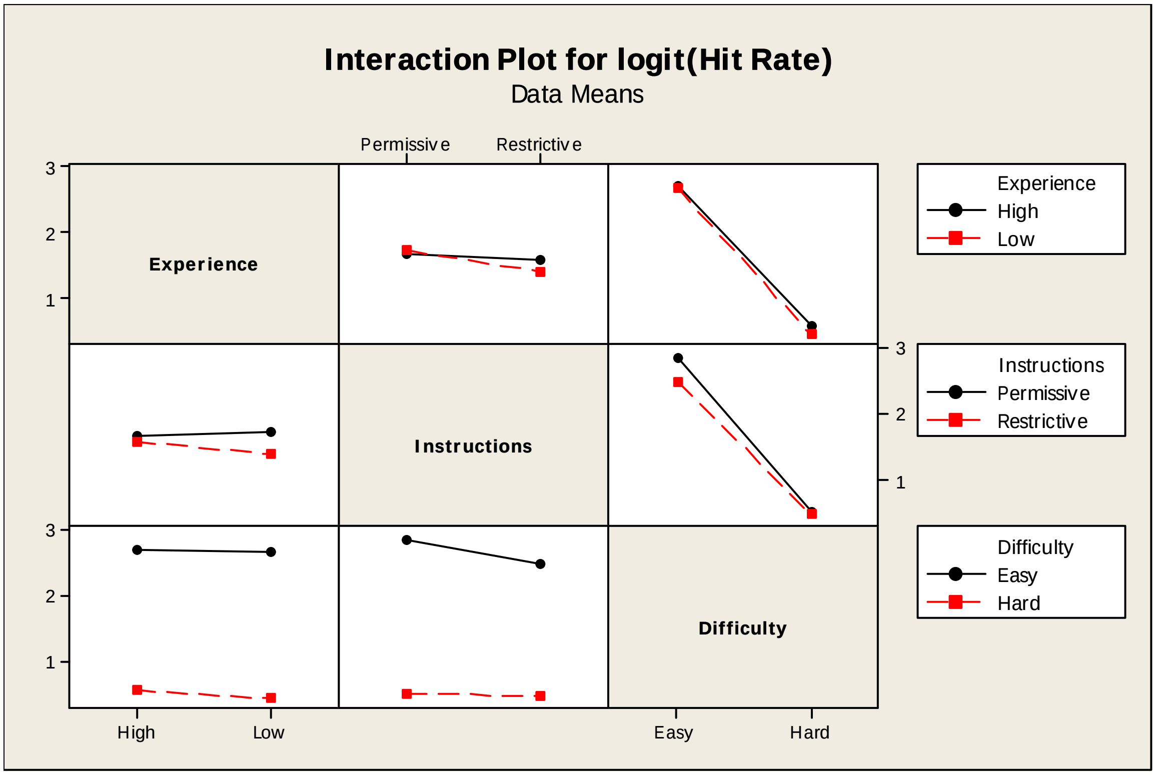

Figure 9. Interaction plot for logit(Hit Rate).

Observer Experience was not shown to significantly affect observers' hit rates. Whereas each combination of Experience, Instructions, and Difficulty had four replications, removing Experience from the plot of individual hit rates (Figure 10) consolidated data across this factor. As a result, there were then essentially eight replications of each combination of scene Difficulty and task Instruction. The means and variances of the raw hit rates appear little changed. The lines connecting the means still have approximately the same slopes and intercepts, indicating that Instructions likewise had no significant affect on hit rates.

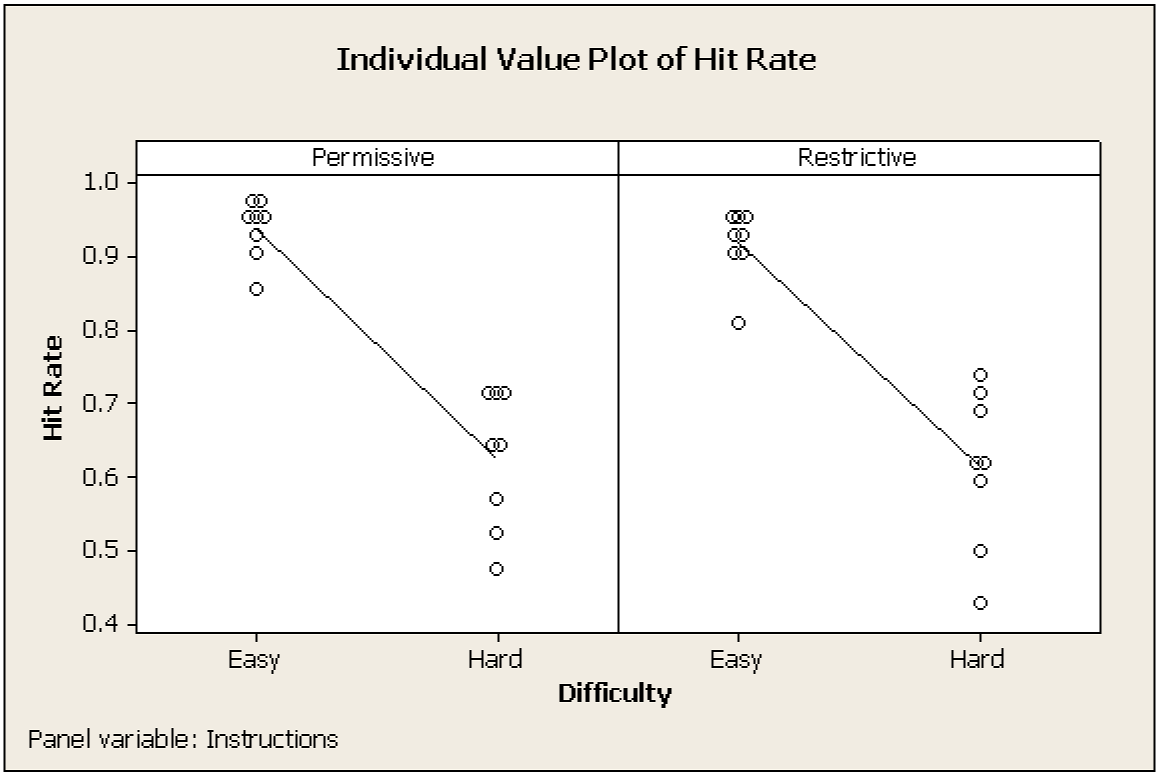

Figure 10. Collapsed individual value plot of hit rates.

C. ANALYSIS OF FALSE ALARM RATES

Similar to the analysis and treatment of raw hit rates, transforming raw false alarm rates with the logit function likewise permitted further analysis by strengthening the assumption of normality and making variances approximately equal. Figure 11 shows the normality probability plots of false alarm rates before and after the transformation. As with hit rates, the transformation brought central data points much closer to the normal line and tail values into the 95% confidence interval.

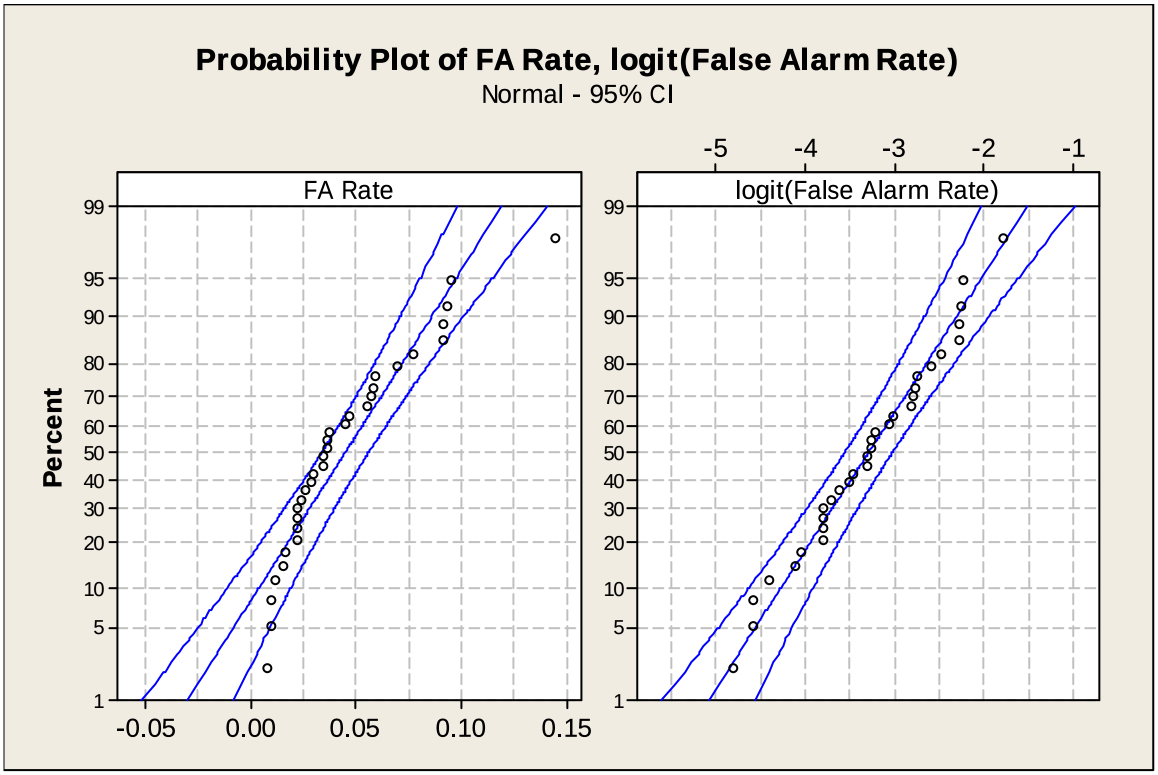

Figure 11. Normal probability plots of False Alarm Rate and logit (False Alarm Rate).

Analysis of the false alarm rates (Table 4) revealed very different findings. The factors that were significant in affecting false alarm rates were Instructions (p-value 0.006), Difficulty (p-value <0.001) and the two-way interaction between Instructions and Difficulty (p-value 0.018).

General Linear Model: logit (False Alarm Rate)

Analysis of Variance for logit(FA Rate)

Source	DF	Seq SS	Adj MS	F	P
Experience	1	0.2105	0.2105	0.38	0.551
Instructions	1	6.3163	6.3163	11.28	**0.006**
Experience*Instructions	1	0.3154	0.3154	0.56	0.467
Participant(Experience Instructions)	12	6.7190	0.5599	5.39	0.003
Difficulty	1	2.5977	2.5977	24.99	**0.000**
Experience*Difficulty	1	0.2581	0.2581	2.48	0.141
Instructions*Difficulty	1	0.7813	0.7813	7.52	**0.018**
Experience*Instructions*Difficulty	1	0.0005	0.0005	0.01	0.944
Error	12	1.2475	0.1040		
Total	31	18.4464			

Table 4. ANOVA of logit(False Alarm Rate)

A plot of the main effects (Figure 12) revealed that task Instructions and scene Difficulty both contributed to false alarm rates. As in the main effects plot for hit rates, observer Experience contributed very little to false alarm rates.

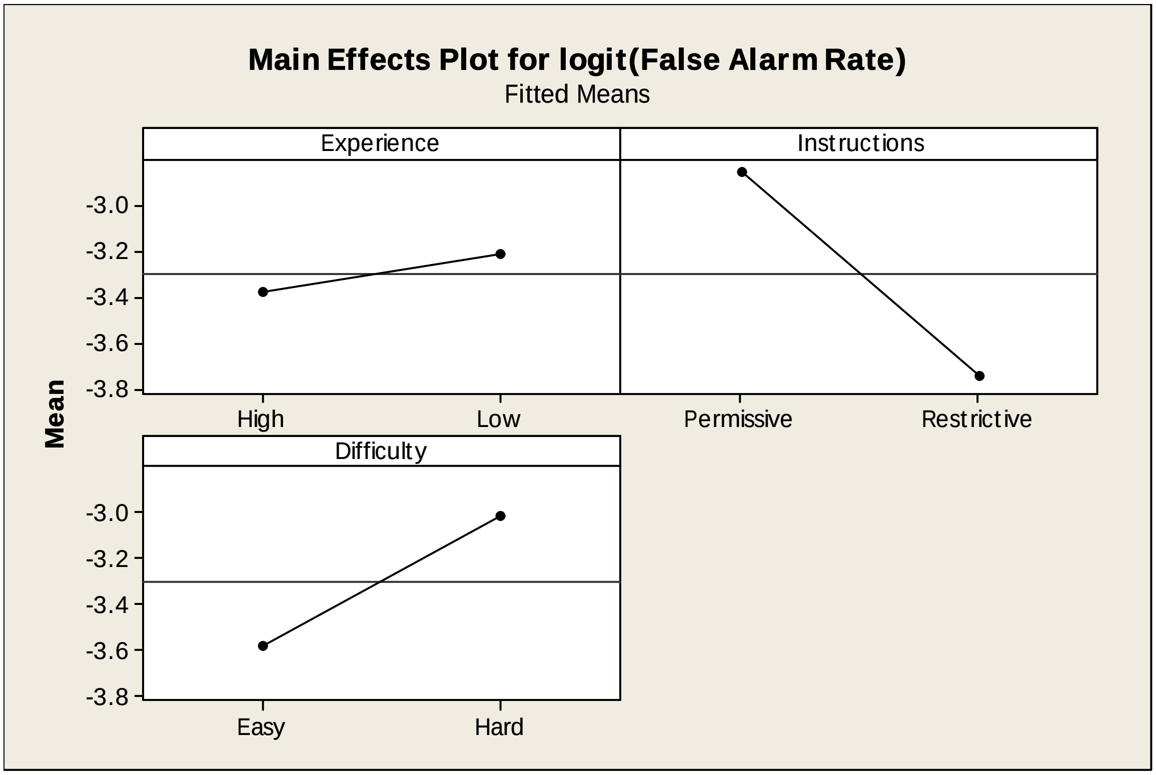

Figure 12. Main effects plot for logit(False Alarm Rate).

With respect to false alarm rates, a plot of the factor interactions (Figure 13) showed that both Instructions and Difficulty contribute to false alarm rates. Additionally, this plot graphically depicts the significant interaction between Instructions and Difficulty.

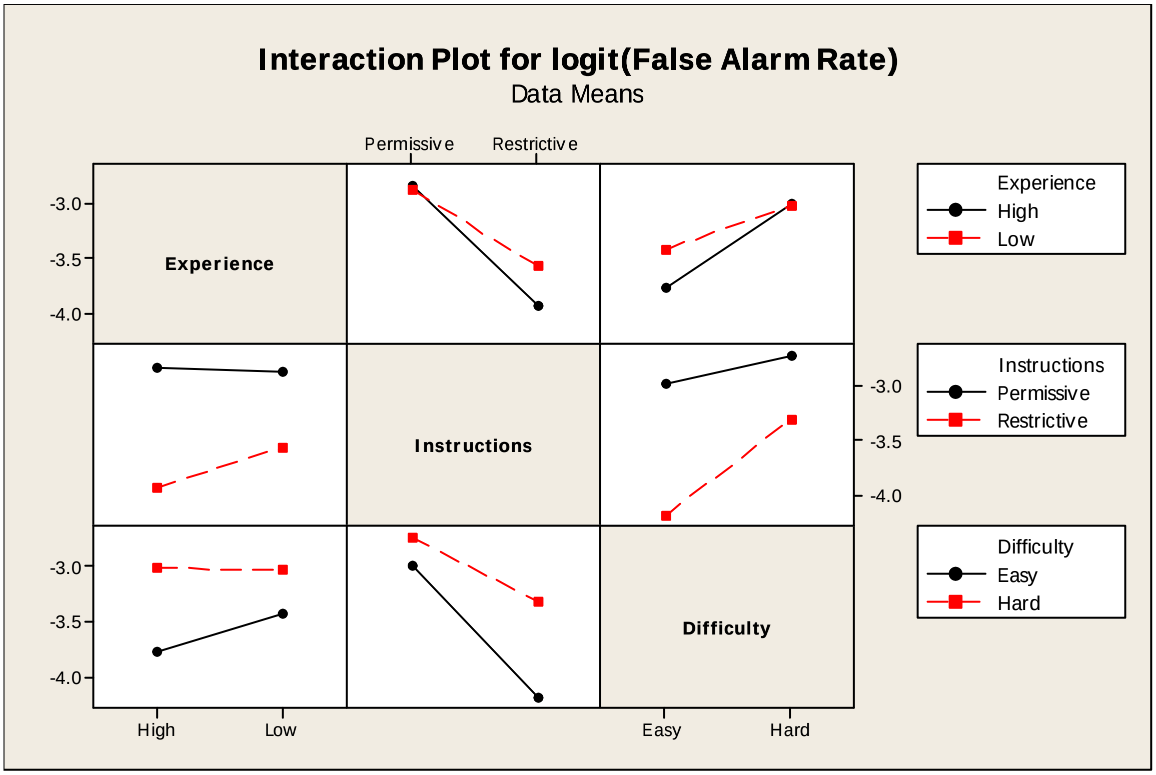

Figure 13. Interaction plot for logit(False Alarm Rate).

Similar to hit rates, observer Experience was not shown to significantly affect false alarm rates. Removing Experience from the plot of individual hit rates again consolidated data so that there were essentially eight replications of each combination of scene Difficulty and task Instruction (Figure 14). The means and variances representing the data from each factor combination continue to widely differ.

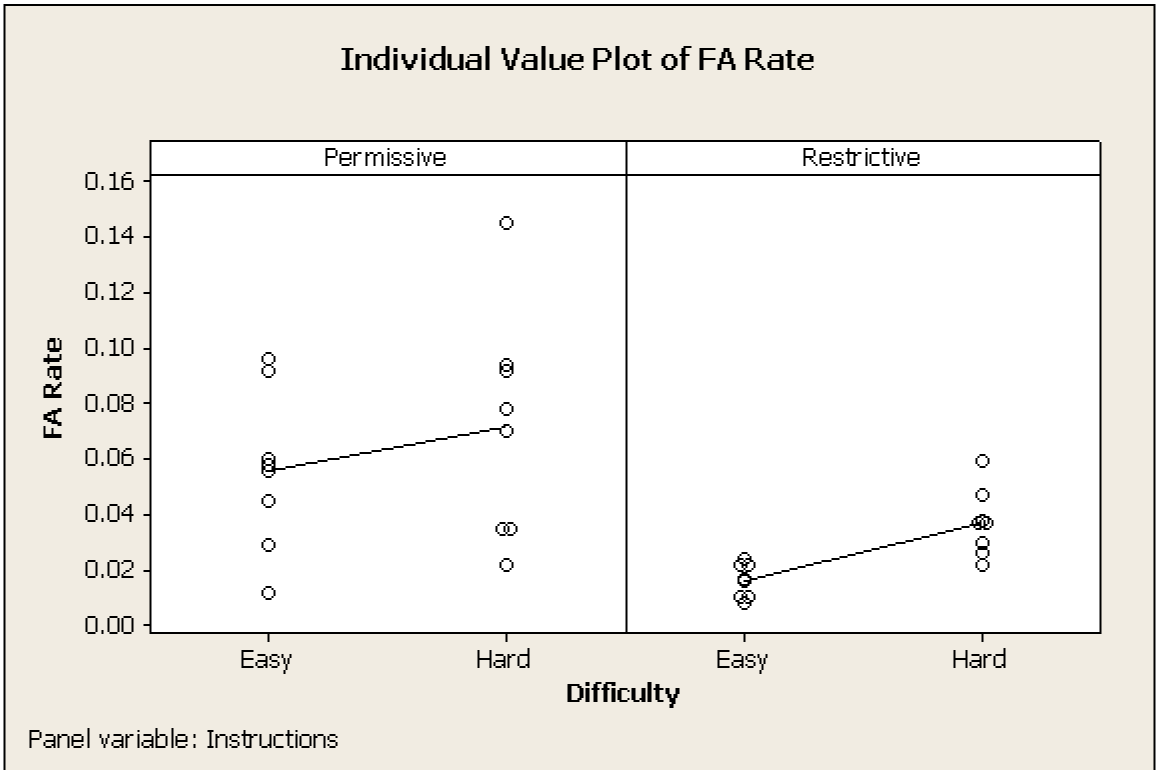

Figure 14. Collapsed individual value plot of false alarm rates.

D. OTHER ANALYSES

Having calculated hit rates and false alarm rates from the experimental data, the framework of Signal Detection Theory enables a more thorough assessment of observer performances by examining sensitivity and bias.

1. Analysis of Observer Sensitivity

As written in Chapter II, an observer's sensitivity is denoted by $\hat{d}'$ and represents the difference between the means of two equal-variance Gaussian curves. It is the difference of the inverse of the standard normal distribution at each observed hit rate from the inverse of the standard normal distribution at each observed false alarm rate. Simply speaking, sensitivity refers to how well an observer can distinguish a signal from the noise. A higher value reflects greater sensitivity. Table 5 displays each observer's sensitivity.

d'		Low Experience				High Experience			
		Permissive		Restrictive		Permissive		Restrictive	
Easy Scenes	Participant Number	32	3.6	31	2.9	21	3.0	22	3.7
		37	3.7	36	3.8	33	3.0	34	3.6
		42	3.2	43	3.6	35	3.9	44	3.5
		46	2.4	45	3.7	52	2.8	51	3.4
	Mean		3.2		3.5		3.2		3.5
	StDev		0.6		0.4		0.5		0.1
Hard Scenes	Participant Number	32	2.4	31	1.4	21	1.9	22	2.1
		37	2.0	36	2.3	33	1.4	34	2.2
		42	2.0	43	2.3	35	2.0	44	2.3
		46	1.4	45	1.8	52	1.6	51	1.9
	Mean		1.9		1.9		1.7		2.1
	StDev		0.4		0.4		0.3		0.2

Table 5. Observer sensitivities.

Similar to previous analysis, the presence of unequal variances among groups is alleviated by transforming the data with the natural logarithm function. Table 6 contains the ANOVA table that shows how, similar to hit rates, the only significant factor was scene Difficulty (p-value <0.001). This was expected, since by design the targets in the hard scenes were much better concealed than the targets in the easy scenes.

General Linear Model: d Prime

Analysis of Variance for d Prime

Source	DF	Seq SS	Seq MS	F	P
Experience	1	0.0002	0.0002	0.00	0.981
Instructions	1	0.5565	0.5565	2.05	0.178
Experience*Instructions	1	0.1028	0.1028	0.38	0.550
Participant(Experience Instructions)	12	3.2566	0.2714	6.17	0.002
Difficulty	1	16.2849	16.2849	369.94	**0.000**
Experience*Difficulty	1	0.0032	0.0032	0.07	0.791
Instructions*Difficulty	1	0.0238	0.0238	0.54	0.477
Experience*Instructions*Difficulty	1	0.0566	0.0566	1.29	0.279
Error	12	0.5282	0.0440		
Total	31	20.8129			

Table 6. ANOVA of observer sensitivity.

The depiction of the main effects plot (Figure 15) displays how scene Difficulty contributes to sensitivity more so than do Instructions or Experience.

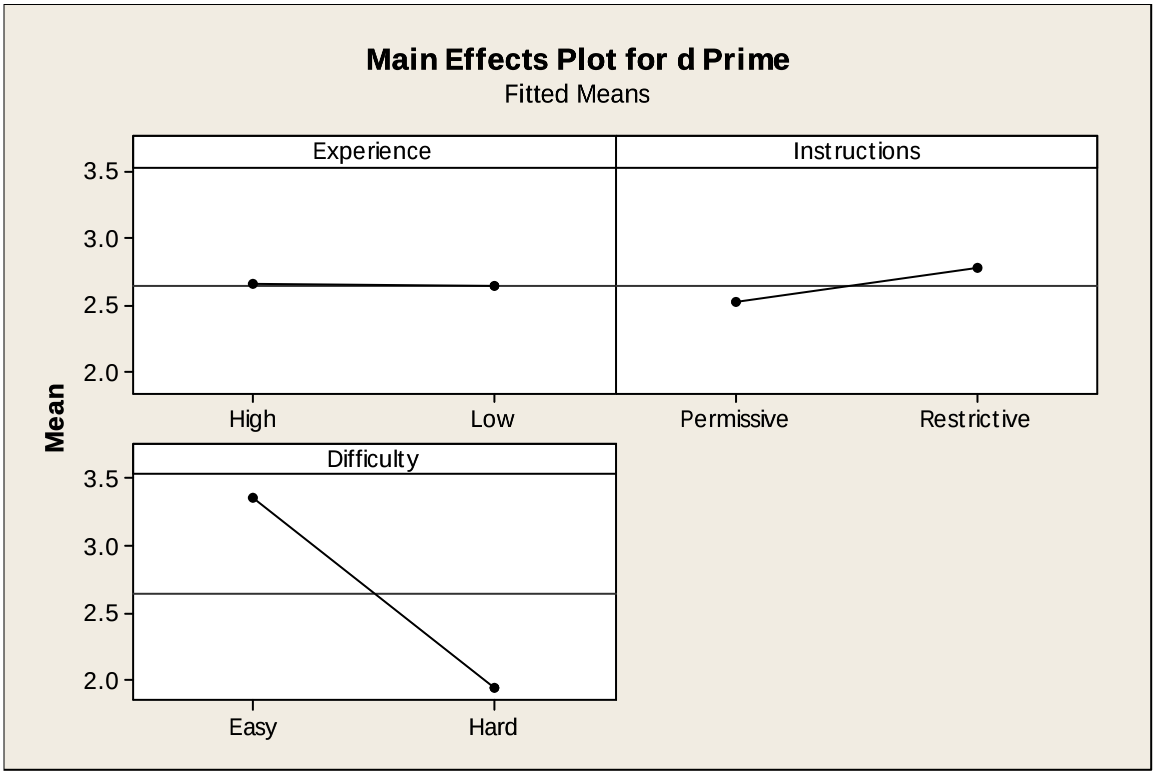

Figure 15. Main effects plot for observer sensitivity.

A plot of all two-way interactions (Figure 16) again shows that Difficulty has the greatest effect, and that there are no significant interactions.

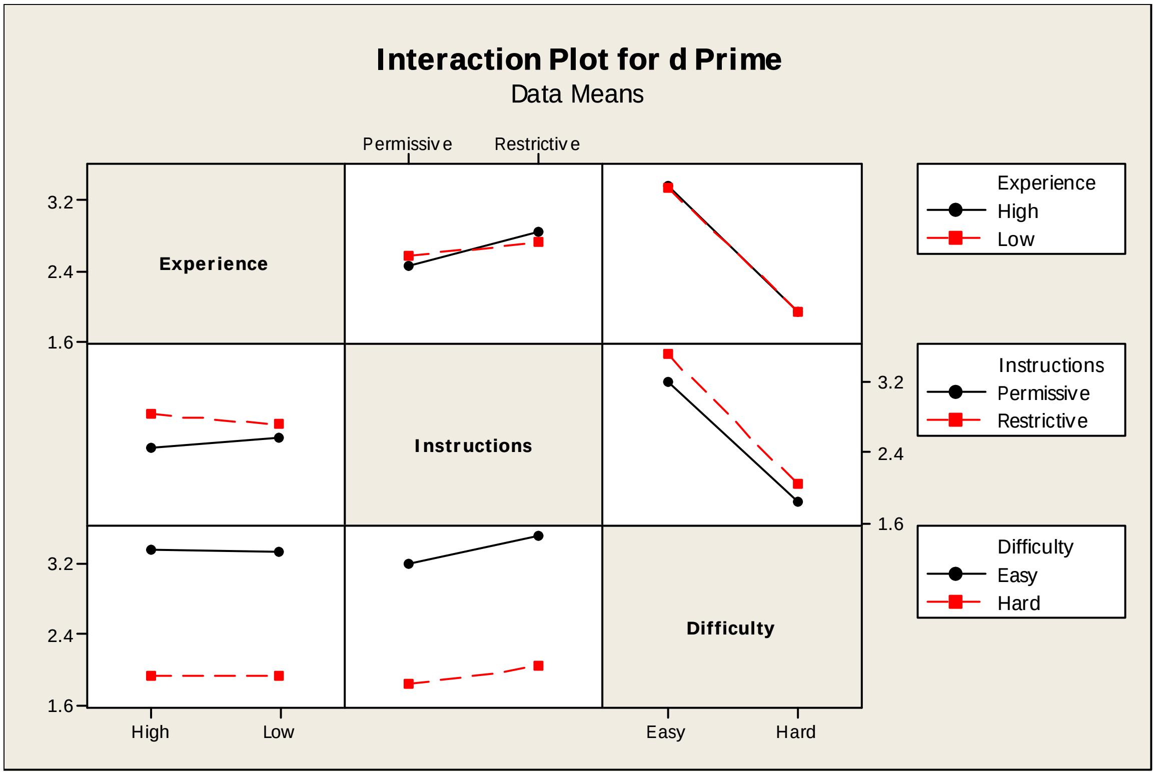

Figure 16. Interaction plot for observer sensitivity.

2. Analysis of Observer Bias

As written in Chapter II, an observer's bias is denoted by β and represents the ratio of the heights of the two standard Gaussian density functions at the observer's criterion. Its present application is as a measure of how much information an observer required before responding with a mouse click. A higher value of bias means that the observer required more information before responding. As a result, such an observer will have accumulated fewer hits and fewer false alarms than an observer with a lower bias. Table 7 displays each participant's bias.

Bias		Low Experience				High Experience			
		Permissive		Restrictive		Permissive		Restrictive	
Easy Scenes	Participant Number	32	-0.7	31	1.6	21	-0.5	22	2.0
		37	-0.5	36	0.9	33	0.2	34	1.8
		42	0.9	43	1.2	35	1.2	44	1.0
		46	0.3	45	0.7	52	-0.7	51	0.1
	Mean		0.0		1.1		0.0		1.2
	StDev		0.8		0.4		0.8		0.9
Hard Scenes	Participant Number	32	1.5	31	1.2	21	0.7	22	1.5
		37	0.8	36	1.2	33	0.9	34	1.9
		42	2.0	43	2.0	35	1.6	44	1.4
		46	0.5	45	1.6	52	0.7	51	0.9
	Mean		1.2		1.5		1.0		1.4
	StDev		0.7		0.4		0.4		0.4

Table 7. Observer biases.

Similar to the findings that resulted from analysis of false alarm rates, analysis of observer biases (Table 8) shows that task Instructions, scene Difficulty, and the interaction between Difficulty and Instructions are the only statistically significant factors (with p-values of 0.019, 0.001, and 0.025, respectively).

General Linear Model: Bias

Analysis of Variance for Bias

Source	DF	Seq SS	Seq MS	F	P
Experience	1	0.0103	0.0103	0.02	0.899
Instructions	1	4.5083	4.5083	7.33	**0.019**
Experience*Instructions	1	0.0323	0.0323	0.05	0.822
Participant(Experience Instructions)	12	7.3760	0.6147	3.26	0.026
Difficulty	1	3.8663	3.8663	20.49	**0.001**
Experience*Difficulty	1	0.1016	0.1016	0.54	0.477
Instructions*Difficulty	1	1.2336	1.2336	6.54	**0.025**
Experience*Instructions*Difficulty	1	0.0021	0.0021	0.01	0.918
Error	12	2.2647	0.1887		
Total	31	19.3951			

Table 8. ANOVA of observer bias.

As shown in the main effects plot (Figure 17), Instructions and Difficulty greatly affect bias while Experience does not. Similarly, the plot of factor interactions (Figure 18) shows great bias differences arising from differing Instructions and Difficulty, and that the only significant interaction was between these two factors.

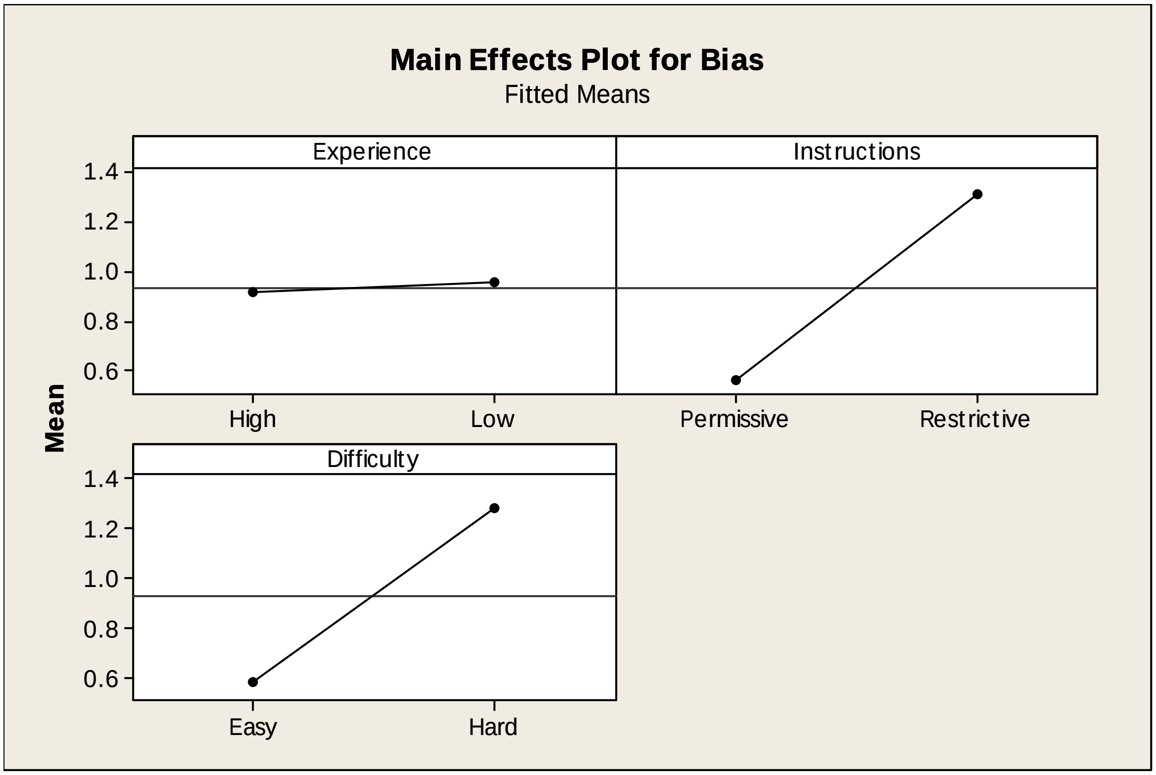

Figure 17. Main effects plot for observer bias.

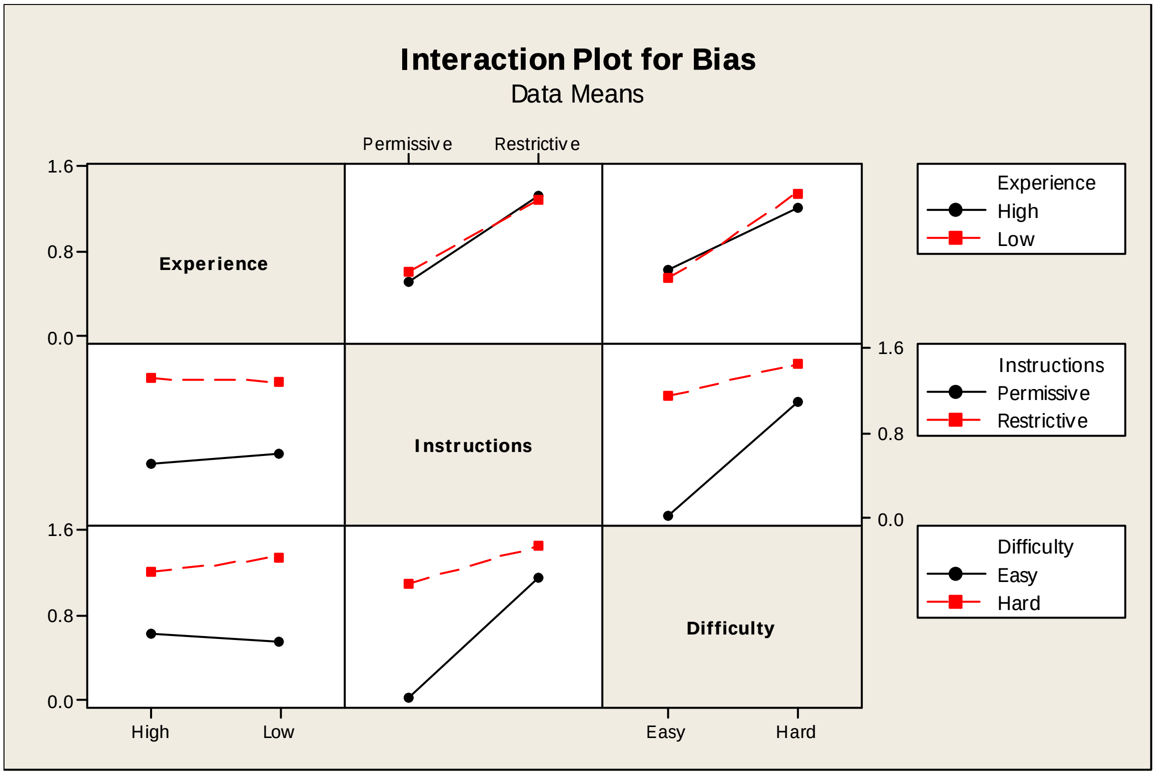

Figure 18. Interaction plot for observer bias.

The plot of factor interactions for bias appears quite similar to the plot of factor interactions for false alarms because the same factors and interactions are at work in much the same ways. Nonetheless, these sorts of plots do not always appear this way; the high degree of similarity in this analysis is a result of the strength the factors and their interaction.

THIS PAGE INTENTIONALLY LEFT BLANK

V. DISCUSSION

A. HYPOTHESES

As discussed in Chapter II, observer performances cannot be appropriately appraised without knowing hit rates and false alarm rates. If false alarms are of little concern, an observer wishing to maximize hits will achieve his goal by simply responding with ‘yes, signal present’ to every trial. Likewise, if hits are of little concern, an observer wishing to minimize false alarms will achieve his goal by simply responding with ‘no, signal absent’ to every trial. As long as there is some uncertainty in the mind of the reasonable observer, increasing hit rates will have the associated effect of increasing the rate of false alarms. Only when considering an observer’s hit rate in tandem with his false alarm rate can one begin to fully understand his target detection performance.

Within the two levels of observer experience and the two levels of task instructions, and between the two levels of scene difficulty, hit rates were fairly uniform. False alarm rates, however, varied greatly within the same experience and instruction levels and between easy and hard scenes. In neither case did experience alone prove to be a significantly contributing factor. Nonetheless, the general uniformity of hit rates in this experiment facilitates further assessment of target detection performance. Analysis of the experimental data leads to rejection of Hypothesis #1 (Experience) and failure to reject Hypothesis #2 (Task Instructions) and Hypothesis #3 (Scene Difficulty).

1. Hypothesis #1 (Experience)

An observer’s experience, as measured by his amount of military target detection training or first-hand combat experience, did not prove to be a significant factor in determining his rate of false alarms. In the ANOVA of logit(False Alarm Rate), shown in Table 4, the p-value associated with the experience factor alone was 0.551. Despite graphical indications of possibly significant interactions, the p-values associated with all two-way and three-way interactions containing the experience factor likewise proved to not be significant. The experimenter thus rejects Hypothesis #1 (Experience): low observer experience did not result in a higher rate of false alarms.

That Experience did not appear significant in this study could be the result of numerous factors. One possibility is that the task simply did not require the unique skills developed while in combat. There are obvious and extreme differences between a desk-side computer 'game' and the life-and-death situations of a combat patrol or convoy. This experiment likely did not evoke the adrenaline, awareness, and intensity that the experienced participants probably exhibited while in combat. Another possibility is that the participants in this study presumably had roughly the same amount of experience searching for 'targets' on a desktop computer. Every participant had been in uniformed service for his or her entire adult life and the nearly universal use of computers in the American military may have further leveled the effects of this factor. Over the years, the participants had refined their abilities to find items of interest on computer screens. Finally, while we naturally expect the most highly trained and experienced observers to be more proficient at a given target detection task than a novice observer, the possibility exists that there is truly no significant difference between such groups.

2. Hypothesis #2 (Task Instructions)

The set of instructions given to observers clearly appeared as a significant factor in influencing false alarm rates. In the ANOVA of logit(False Alarm Rate), the p-value associated with the Instructions factor alone was 0.006. The steep line connecting the fitted means associated with 'Permissive' and 'Restrictive' instructions in the graphical depiction of main effects (Figure 12) highlights the profound impact of this factor. Contributing to the great difference is the significant interaction between Instructions and Difficulty (p-value 0.018). The experimenter thus fails to reject Hypothesis #2 (Task Instructions): instructions are indeed shown to have great effect on the rate of false alarms.

The instructions given to participants were designed to elicit greatly differing responses while still permitting the possibility of identical target-detection performance. Experimenters took great care to develop the written instructions and the training scenario so that participants would sufficiently understand and apply the given instructions while not unduly constraining natural behavior. The ‘Permissive’ instructions emphasized hits while making no mention of false alarms. (Indeed, many participants asked questions about or made mention of the absence of instructions concerning false alarms. In such instances, the experimenter simply referred them to the appropriate portion of the written instructions.) After the first of three practice sessions, most participants given ‘Permissive’ instructions quickly realized how liberal mouse-clicking increased the number of hit targets. In order to keep from unequally biasing participants, experimenters made no mention of a participant’s false alarms accumulated during the practice sessions.

The ‘Restrictive’ instructions also emphasized hits but required participants to be much more judicious in their mouse-click behavior. Whereas participants given ‘Permissive’ instructions could click upon a suspected target without a second thought, participants given ‘Restrictive’ instructions would seemingly maximize the number of hits by requiring more certain information about a suspected target’s identity before responding. These participants thus had to be much more careful as they searched each scene and respond to targets that provided an amount of information that exceeded their own thresholds of certainty.

3. Hypothesis #3 (Scene Difficulty)

The primary purpose of the pilot study was to gain insight into the factors that would enable development of a set of stimuli resulting in a large distinction between observer performances. Experimenters then developed the main experiment’s set of target detection scenes with the expressed purpose of making ‘easy’ and ‘hard’ scenes. It therefore comes as no surprise that Difficulty significantly influenced false alarms (p-value <0.001, as shown in Table 4). The experimenter thus fails to reject Hypothesis #3 (Scene Difficulty): indeed, more difficult scenes generate higher false alarm rates.

B. ADDITIONAL FINDINGS

1. Factor Interactions

A review of the mean false alarm rates (Table 2) shows that, regardless of whether the observer is experienced or inexperienced, false alarm rates are lowest when the scene difficulty is easy and when the observer is given restrictive instructions. Conversely, the highest mean false alarm rates resulted when observers given permissive instructions examined hard scenes.

In addition to the significant main effects of Instructions and Difficulty, analysis shows that the interaction between these two factors is also significant (p-value 0.018). This highly significant interaction between Instructions and Difficulty appeared in the difference in magnitude within scene Difficulty and between task Instructions. That is, the effects of permissive and restrictive instructions on easy and hard scenes were more than just additive. On one hand, a typical observer given permissive instructions had only a slightly higher false alarm rate on hard scenes than on easy scenes. On the other hand, a typical observer given restrictive instructions had a much higher false alarm rate on hard scenes than on easy scenes. It is important to note that all such differences were the result of numerous responses to the same stimuli. When presented the same set of target detection scenes, observers given permissive instructions generated more false alarms. Similarly, regardless of whether the scene was easy or hard, the typical observer given permissive instructions responded with more false alarms than the observer given restrictive instructions.

On the interactions plot of logit (False Alarm Rate) in Figure 19, the mean logit(False Alarm Rate) is approximately -3.0 for Easy/Permissive and -4.2 for Easy/Restrictive. Converting these factors back to standard false alarm rates by applying the appropriate exponential function reveals mean false alarm rates of 0.047 (Easy/Permissive) and 0.015 (Easy/Restrictive), for an Instruction-induced difference of 0.032. On the same plot, the mean logit(False Alarm Rate) is -2.7 for the combination of Hard/Permissive and -3.4 for Hard/Restrictive. Re-converting these values as above yields standard mean false alarm rates of 0.060 and 0.036, respectively, for an

Instruction-induced difference of 0.024. Comparing these differences (0.032 for Easy and 0.024 for Hard) reveals a 33% difference in mean false alarm rates due to the interaction between Difficulty and Instruction.

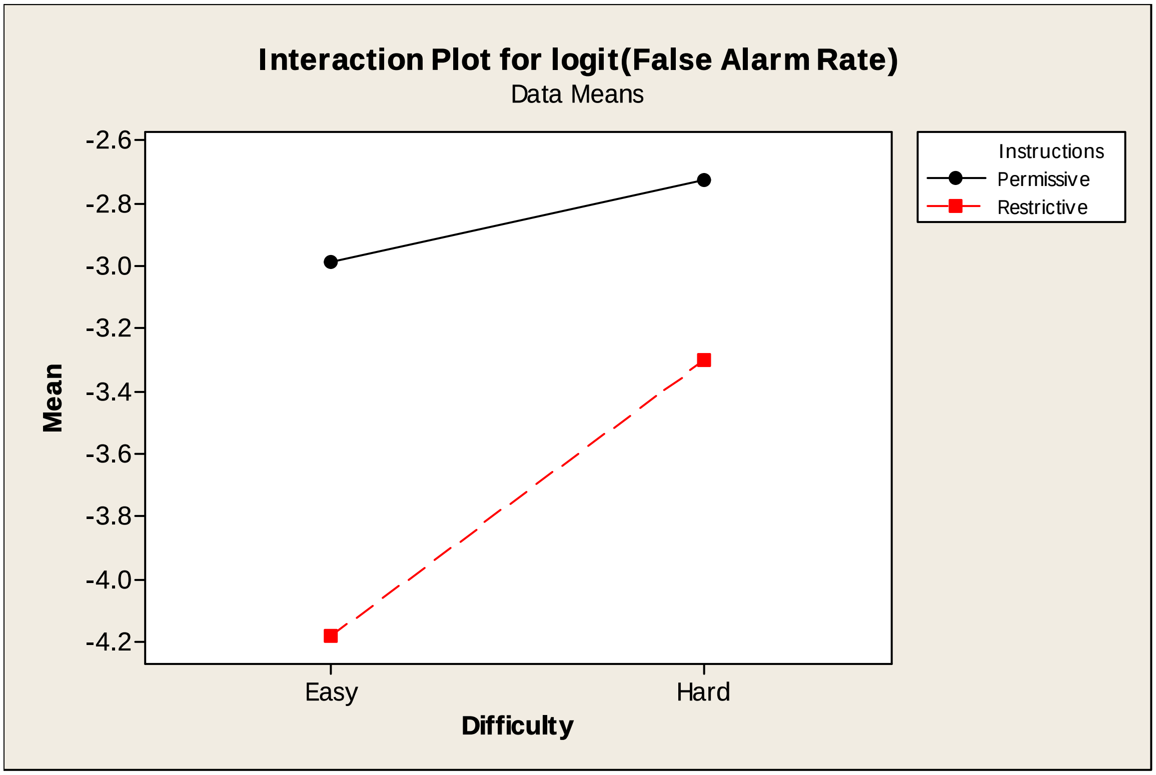

Figure 19. Interaction plot of instructions and difficulty for logit(False Alarm Rate).

2. Performance Comparisons

As discussed earlier, hits and hit rates were higher on easy scenes while false alarms and false alarm rates were higher on hard scenes. Although these broad descriptions confirm intuition, they nonetheless conceal important distinctions. Since a hit rate is the quotient of hits divided by the sum of hits and misses, reporting a hit rate alone conceals the effects of each factor. That is, knowing the numerator and denominator provide important information. The same is true for understanding false alarm rates.

With regard to easy scenes, the average number of hits by observers given permissive instructions (39.4) is nearly identical to that by observers given restrictive instructions (38.5). The same is true with regard to hard scenes: differing instructions result in little difference between average numbers of hits (26.2 and 25.8, respectively). As a result, the ratios of hits across easy and hard scenes are both approximately equal to 1.0, as shown in the final column of Table 9. Figure 20 graphically displays the average number of hits for each combination of Difficulty and Instruction, with error bars showing one standard deviation.

		Instructions		
		Permissive	**Restrictive**	**Ratio**
Easy Scenes	Hits	39.4 (1.7)	38.5 (2.0)	1.02
	Misses	2.6 (1.7)	3.5 (2.0)	0.74
	Hit Rate	0.938	0.917	1.02
		Permissive	**Restrictive**	**Ratio**
Hard Scenes	Hits	26.2 (3.9)	25.8 (4.5)	1.02
	Misses	15.8 (3.9)	16.2 (4.5)	0.98
	Hit Rate	0.624	0.614	1.02

Table 9. Average hits and misses (standard deviations in parentheses).

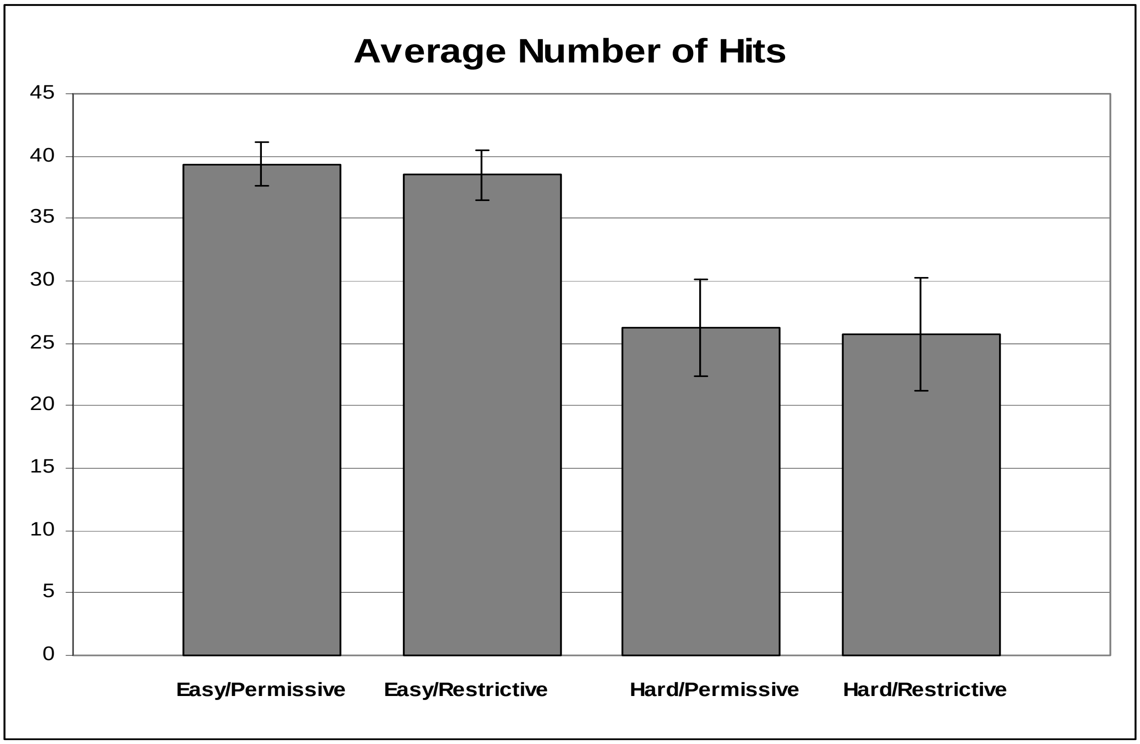

Figure 20. Average numbers of hits for each difficulty and instruction.

The number of hits on easy scenes is clearly higher than the number of hits on hard scenes. The form of Instructions, whether permissive or restrictive, did not significantly affect the number of hits in either type of scene Difficulty. Within each level of Difficulty, the ratio of hits resulting from permissive instructions to hits resulting from restrictive instructions is approximately one (Figure 21).

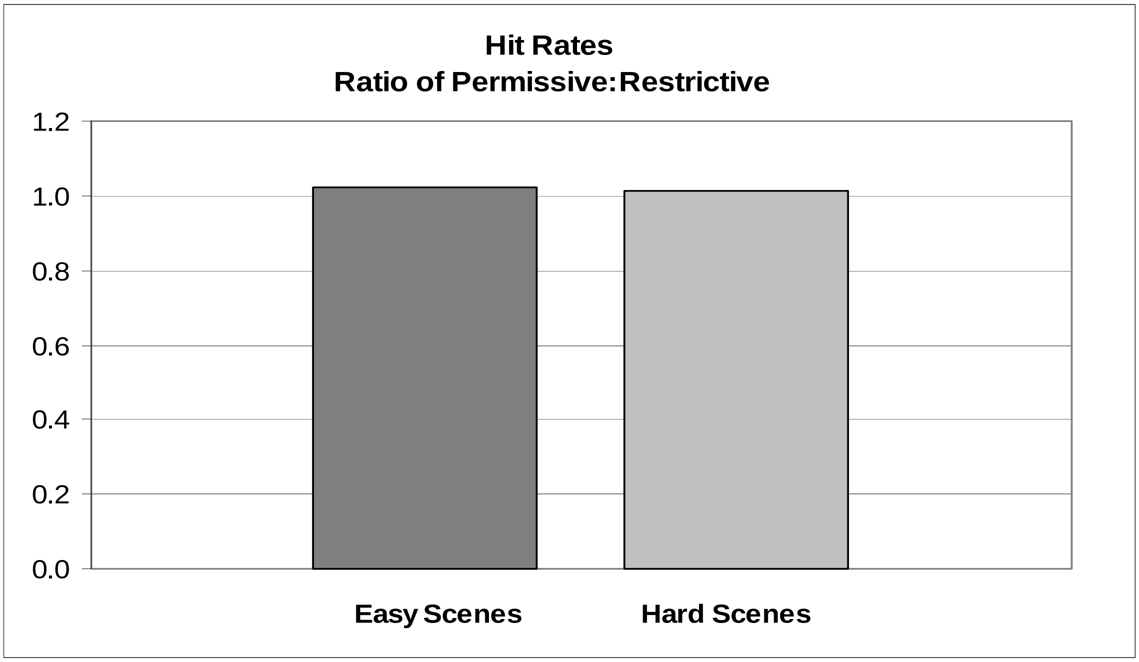

Figure 21. Effects of instructions on hit rates in easy and hard scenes.

In terms of false alarms, the data tell a very different story. With regard to easy scenes, the average number of false alarms by observers given permissive instructions (10.4) is nearly 3.5 times that by observers given restrictive instructions (3.0). Their average numbers of correct rejections only slightly differed, resulting in a false alarm rate by permissively instructed observers that is more than three times higher than the false alarm rate of restricted observers. Table 10 displays the average numbers of false alarms and correct rejections. Figure 22 graphically displays the average number of false alarms for each combination of Difficulty and Instruction, with error bars showing one standard deviation.

		Instructions		
		Permissive	**Restrictive**	**Ratio**
Easy Scenes	False Alarms	10.4 (6.0)	3.0 (1.5)	3.46
	Correct Rejections	156.4 (30.8)	151.4 (40.1)	1.03
	False Alarm Rate	0.062	0.019	3.20
		Permissive	**Restrictive**	**Ratio**
Hard Scenes	False Alarms	13.6 (7.4)	7.0 (2.7)	1.95
	Correct Rejections	174.0 (39.2)	162.4 (38.4)	1.07
	False Alarm Rate	0.073	0.041	1.76

Table 10. Average false alarms and correct rejections (standard deviations in parentheses).

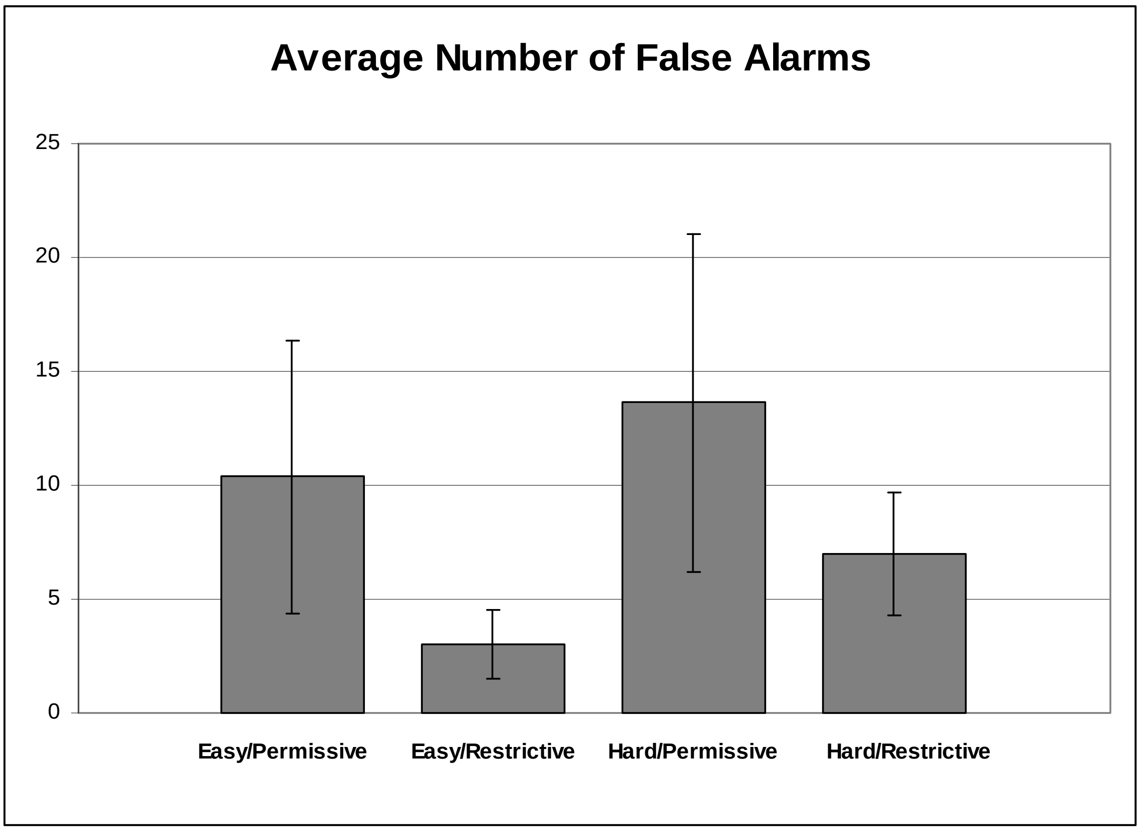

Figure 22. Average numbers of false alarms for each difficulty and instruction.

To a lesser extent, the same is true for observers of hard scenes. Those given permissive instructions averaged 13.6 false alarms, while those given restrictive instructions averaged only 7.0 false alarms. The difference between the average numbers of correct rejections also increased compared to easy scenes, indicating that those given permissive instructions tended to search more thoroughly. While the average number of false alarms thus increased compared to easy scenes, the ratio of average false alarms decreased: from 3.2 for easy scenes to 1.8 for hard scenes. See Figure 23.

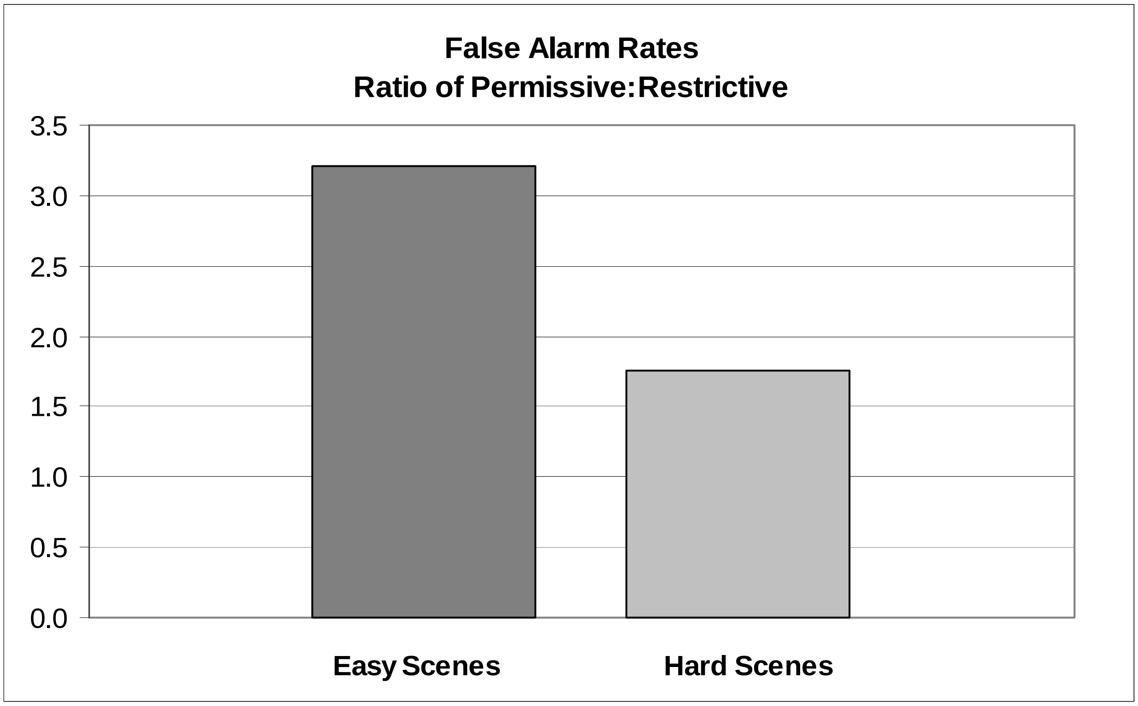

Figure 23. Effects of instructions on false alarm rates in easy and hard scenes.

What caused these differences? Why were there such big differences in the numbers of false alarms among differently instructed observers? And why was the difference bigger among easy scenes than among hard scenes? The answers to these questions lie in the calculation of false alarm rates: numerators and denominators.

Observers given restrictive instructions implicitly required a higher degree of certainty before deciding whether an object was a target. On easy and hard scenes, they hit just as many targets as their counterparts, but were far less likely to incorrectly click upon a non-target. Hard scenes merely appeared to them as scenes containing few targets. They generated a low number of false alarms because they were not inclined to 'waste' their few precious available shots.

Since observers given permissive instructions were not penalized for incorrect responses (but were nonetheless restricted from taking “wild shots”), they required a low degree of certainty before clicking upon a suspected target. On easy and hard scenes, their average numbers of hits were very comparable to observers given the opposite instructions. On hard scenes, their average false alarm rates were nearly twice as high those given restrictive instructions, and more than three times as high on easy scenes. Since these observers were permitted to click more liberally, they were very likely to approach the upper limit of targets within a scene (identified in the instructions as “up to six targets per scene”). Having viewed some scenes with no readily apparent targets, these observers viewed subsequent easy scenes in much the same manner: wanting to maximize the number of hits, they correctly identified the obvious targets but nonetheless required very low certainty before clicking upon non-targets. Observers given permissive instructions were thus much more willing to click upon non-targets in easy scenes. Since most scenes contained fewer than six targets, they were very likely to accrue false alarms.

3. Bias and Sensitivity

Continuing the analysis by calculating and analyzing each participant’s measured values of sensitivity (d’) and bias (β) helped to more fully understand the target detection performance of the experiment’s participants. Calculation of d’ used both an observer’s sample hit rate and false alarm rate. It represents an observer’s ability to accurately distinguish ‘signal plus noise’ events from ‘noise alone’ events. As in the analysis of the hit rates, the most important factor in determining an observer’s sensitivity was scene Difficulty (p-value <0.001). In this study, an observer’s sensitivity was thus affected only by the saliency of the target within the scene. Neither combat experience nor instructions had any apparent bearing on how well a participant detected true targets within the scenes. Given this study’s relatively homogeneous sample population (well-educated mid-career active duty U.S. military officers) and the physical characteristics of the observation task, it was unsurprising that the participants exhibited very similar sensitivities to targets.

Finally, analysis of bias revealed the primary factors underlying an observer's placement of criterion. The factors significantly affecting bias in this study were Instructions (p-value 0.019), Difficulty (p value 0.001), and the two-way interaction between Instruction and Difficulty (p-value 0.025). Similar to the findings of significant factors driving false alarm rates, the placement of an observer's criterion (that is, the balancing of his desires to correctly hit targets and reject non-targets) was dependent on the level of scene Difficulty, the task Instructions, and the interaction between Difficulty and Instruction. The combat Experience of the observer had no significant bearing on resultant criterion.

Generally, effective models are those that include only the most essential factors. Higher-order interactions rarely have important places in such models. It was unsurprising that the three-way interaction between Experience, Instruction, and Scene Difficulty was not important to predicting hit rate, false alarm rate, sensitivity, or bias. The significant two-way interaction between Instruction and Difficulty in predicting false alarm rates and observer bias was nonetheless an unexpected but highly important outcome.

C. IMPLICATIONS

False alarms will occur as long as imperfect detectors seek to separate signals from noise. In order to accurately represent human performance, then, developers of combat models must no longer model only the physical properties of target detection. They must consider and account for the observer's propensity to falsely detect targets. In doing so, they must develop methods to incorporate the observer's instructions and an assessment of the 'difficulty' of finding the target in the scene.

As of this writing, American troops have been in Iraq for just over five years. Soldiers and Marines have conducted thousands of convoys and foot patrols and have accumulated a wealth of combat experience that the services have not seen in more than a generation. Given that so many of the current combat models were conceived and developed during and immediately following the Cold War, there is some doubt as to whether these combat models reflect the growing body of 'institutional knowledge.' Having determined that observer Experience does not significantly affect false alarm rates, the results of this experiment may allay such doubts.

The critical distinctive of Instructions is *how* the observer is to accomplish the assigned task. The spectrum of observer's instructions can include such sources as: cultural perspectives, command climate, unit morale, mission statements, rules of engagement, ammunition supply, stress, and emotion. Although quantifying and modeling these parameters may be daunting tasks, it is clear that it is important.

Quantitatively assessing the Difficulty of the observer's scene appears much more feasible. Methods and techniques exist that enable modelers to account for the target's color, shape, texture, and brightness, and also the surrounding area's color texture, brightness, contrast, and clutter. Modeling and rendering the physical properties of one entity seeing another is a task the combat modelers are tackling now.

Modelers must continue to develop precise qualitative and quantitative methods to assess the difficulty of an entity's scene and establish a means of integrating the observer's instructions. In addition to these steps, modelers must also connect these two factors such that they capture their apparent interactive effects. Failure to do so is an implicit acceptance of lower-resolution models.

D. CONNECTIONS TO LITERATURE REVIEW

The importance of Difficulty and Instructions in false positive detections is a finding that significantly affects existing combat models and the application of current military doctrine and previous eye tracking studies.

1. Combat Models

The apparent significance of Instructions reveals weaknesses in the ways combat models such as ACQUIRE and CASTFOREM account for false positive detections. The ACQUIRE algorithm emphasizes the physical aspects of target detection and therefore neglects a critical aspect of representing false positive detections. By ignoring

Instructions and the Difficulty-Instructions interaction, it essentially accounts for only a third of the factors contributing to false alarms. The results of this research provide good reason for modelers to gather information that will appropriately account for such transient factors as commander's intent and rules of engagement.

This research also shows that CASTFOREM's approach to false positive detections is inappropriate. It relies only on distributions of true targets when the rate of false positive detections requires information concerning observers and targets alike. By inserting additional entities into the scenario and internally labeling them as false targets, CASTFOREM implicitly rejects this experiment's findings. The results of this experiment show that CASTFOREM's method of presenting true targets and treating them as false targets is inaccurate.

2. Doctrine

Military doctrinal manuals prescribe a two-step search technique that is influenced by the Soldier's bias. They suggest that false positive detections tend to occur in initial searches and that continued searching reduces the number of missed targets. The first step encourages hitting a high number of targets and missing very few of them. The second step requires more time and encourages observers to avoid missing targets while still keeping low the number of false alarms. Detecting and responding to a target at any point in the search process forces the observer to begin the process anew.

The first step (hasty search) coincides with a Soldier's natural desire to quickly detect the enemy. An observer using this technique requires low certainty about a suspected target's identity, thereby permitting a high number of false positive detections. The recommendation of hasty searches implies that the costs of missing targets are much higher than the costs of false positive detections. These are the fundamental characteristics of a low bias condition regardless of the observer's detection capability.

The task in this experiment reflected the initial, hasty search of an area. Limiting observation time to 15 seconds forced observers to rapidly and actively search for targets in the entire scene. The time limit generally prohibited observers from transitioning to the slower and more deliberate search method. According to the results of this experiment, Soldiers given low bias instructions will be even more apt to immediately and incorrectly respond to non-targets.

3. Signal Detection Theory

Signal Detection Theory must allow for an expansion of terms. In experiments of the classic Yes-No or Two-Alternative Forced-Choice Tasks, the current four response descriptors (hit, miss, correct rejection, and false alarm) suffice. With the advent of high-fidelity eye-tracking technology the original four descriptors insufficiently describe the range of responses in multiple-target scenarios. More specifically, there must be greater resolution between types of misses. There must be a distinction between 1) simply not looking at a target (due to excellent camouflage, neglect, or even apathy) and 2) looking at a target but choosing not to declare it a target. The former might be called an 'Oversight,' while the latter might be called an 'Incorrect Rejection.' Analysts must review these classifications in terms of their implications on the current definitions of hit rates, false alarm rates, and ROC curves. To continue this study in light of this expanded set of observer responses, an experimenter may count the number of times a participant looked at a target but chose not to click on it. All such events would be called 'Incorrect Rejections.' The experimenter would then examine if these incorrect rejections are the result of observer experience, instructional bias, and scene difficulty.

4. Eye Tracking Research

This experiment was greatly influenced by the works of many excellent target detection and eye tracking researchers. Their findings influenced the choice of participants and equipment, the development of instructions, and the experimental procedures. Although previous researchers analyzed and reported false alarms and false alarm rates only as incidental to studies emphasizing hits and hit rates, this research generally confirms their conclusions.

The choice of participants and equipment reflect the nature and purpose of the study. Since the search of enemy personnel lies squarely within the military domain, active duty military personnel were naturally chosen for this experiment. The Naval Postgraduate School provided a deep pool of such participants. Participants with broad ranges of experiences permitted arriving at the conclusion that Experience significantly affected neither hit rates nor false alarm rates. Without a breadth of training and combat backgrounds, the experimenter would not have been able to make this distinction.

The instructions given to participants clearly affected target detection performance. As noted in Chapter II, Ozkaptan (1979) and Pavel (1987) both recognized that instructions given to participants greatly influence subsequent performance. More specifically, Yeh (1998) concluded that observer expectancies improve performance when targets are present, but hinder performance when targets are absent. Yeh's conclusion sheds additional light on the significant effect of the Difficulty-Instructions interaction. Finally, See (1997) found that observer bias decreased as the number of targets increased. Since most of their experiments presumably did not have access to eye tracking technology of sufficient fidelity, they were unable to track and record the 'correct rejections' of participants. Accordingly, these researchers focused their research and analysis primarily on hits and hit rates. Nonetheless, their general observations and this experiment confirm that their conclusions apply to false alarms and false alarm rates, too.

VI. CONCLUSION AND RECOMMENDATIONS

A. IMPLICATIONS FOR COMBAT MODELS

Target detection algorithms emphasize the physical aspects of detection in favor of accurately representing hit rates. In doing so, these models poorly represent actual false alarms and largely neglect an observer's bias. This research confirmed that the current approach adequately models hits and hit rates. In terms of false alarms and false alarm rates, however, this research showed that simply assessing the scene's difficulty is insufficient. The set of observer instructions was shown to also be a significant factor. Furthermore, there was significant interaction between a scene's difficulty and the observer's instructions.

Increasing the complexity of target detection algorithms with additional factors may prove to be computationally expensive. If model developers and users determine that the accurate representation of false alarms remains unimportant, then there is little to be gained by incorporating such complicating factors into future combat models. Despite this seemingly high hurdle, the increasing complexity of modern warfare and the growing reliance upon modeling and simulation for realistic training demands that such models accurately represent human behavior.

Developers of combat models must continue to improve target detection algorithms. They must design and create entities that have hit rates and false alarm rates that closely approximate the hit rates and false alarm rates of human observers. To maximize training effectiveness, combat models and simulations must incorporate the significant effects of observer instructions and bias.

B. AREAS OF FUTURE RESEARCH

The purpose of this experiment was to explore the causes of false alarms in the unaided visual search for human targets in an urban combat environment. The experiment's ultimate goal is the improvement of combat models and simulations. While the findings of this research proved significant, there is still much to be learned.

Analysis of this experiment revealed no substantial difference in target detection performance among officers with extensive infantry experience versus other military officers. In assigning observers to either the 'high experience' or 'low experience' groups, there was no distinction between target detection training and target detection experience. As a result, this study included no analysis of their differences. To thoroughly assess this apparent indistinguishability, future experimenters may wish to study the target detection behavior of more diverse groups of subjects. Repeating this study with infantrymen having immediately returned from deployments featuring frequent urban combat versus ROTC or academy cadets will provide a more robust comparison.

Future experimenters may wish to obtain more specific data about individual targets and scenes. This experiment included neither analysis of the time to detect individual targets nor of their sequence of detection within each scene. Studying these dependent variables may shed even more light upon the target detection sequence.

Finally, those conducting future research may seek to develop a means to increase the realism of the target detection scene. Instead of a single computer screen they may use multiple screens to create a 180 degree scan. Instead of a mouse pointer they may use a laser-equipped rifle. Instead of all having all targets in a scene appear either well-exposed or well-hidden they may devise a way to mix such targets without compromising the ability to capture an observer's correct rejections.

APPENDIX A: PARTICIPANT SCREENING SURVEY

TARGET DETECTION IN COMBAT MODELS

Thank you for your willingness to participate in this research project. Please answer the following questions as precisely as possible.

1. Name: ________________________ Email: _____________________@nps.edu

2. Citizenship: United States Other:______________________________________

3. Branch of service: Army Navy Marine Air Force Other: ________________

4. How experienced, trained, or qualified are you in the following Urban Operations (UO) areas (circle as appropriate):

	Low			Moderate			High		
a. Close Quarters Battle	*	*	*	*	*	*	*	*	*
b. Leading Convoys	*	*	*	*	*	*	*	*	*
c. Convoy Security	*	*	*	*	*	*	*	*	*

5. Have you deployed to Iraq or Afghanistan since April 2003?
   No Yes Return date: _____________________

6. If you have deployed to Iraq or Afghanistan, approximately how many times did you

   a. conduct a patrol on foot? ____

   b. conduct a vehicle-mounted patrol (or convoy)? ______________________

7. Briefly describe any of the following specialized **target-detection training** or **experiences** you may have received (such as sniper, scout/reconnaissance, shoot/don't shoot exercises)

   ____

   ____

8. Do you have a visual correction prescription? Yes No
   a. If Yes, do you wear glasses? Yes No
   b. If Yes, do you wear contact lenses? Yes No
   c. If No, have you received visual corrective surgery? Yes No

9. How often do you usually play video/computer games featuring first-person shooters?
    a. Almost every day
    b. Once or twice per week
    c. Once or twice per month
    d. Less often

We will review your responses and contact you to arrange a time for completion of the target detection scenario.

In the 24 hours prior to you participation, please do not depart from your usual sleep routine and use of caffeine, tobacco, and/or alcohol.

If you have a visual correction prescription, please wear your glasses to the experimental session. The wearing of contact lenses will inhibit data collection.

Thank you!

APPENDIX B: INFORMED CONSENT WAIVER

Naval Postgraduate School

Informed Consent Form

Introduction. I understand that have been invited to participate in a study entitled Target Detections in Combat Models being conducted by the Operations Research Department and MOVES Institute of the Naval Postgraduate School.

Procedures. I understand that the purpose of this research is to improve the target detection characteristics employed by current and future combat models. I understand that I will be shown several scenes from a fictitious combat environment and that these scenes may contain varying numbers of human-like targets. I understand that I will use typical office computer equipment to indicate target detection and that approximately 45 minutes may be required for set-up, familiarization, scene presentation, and questionnaires.

Risks and Benefits. I understand that this project does not involve greater than minimal risk and involves no known reasonably foreseeable risks or hazards greater than those encountered in everyday life. I have also been informed of any benefits to myself or to others that may reasonably be expected as a result of this research.

Compensation. I understand that no tangible compensation will be given. I understand that a copy of the research results will be available in the student thesis section of the Dudley Knox Library at the conclusion of the experiment.

Confidentiality & Privacy Act. I understand that all records of this study will be kept confidential and that my privacy will be safeguarded. No information will be publicly accessible which could identify me as a participant. I will be identified only as a code number on all research forms/data bases. My name on any signed document will not be paired with my code number in order to protect my identity. I understand that records of my participation will be maintained by NPS for three years, after which they will be destroyed.

Voluntary Nature of the Study. I understand that my participation is strictly voluntary, and if I agree to participate, I am free to withdraw at any time without prejudice.

Points of Contact. I understand that if I have any questions or comments regarding this project upon the completion of my participation, I should contact the Principal Investigator, Dr. Lawrence G. Shattuck, 656-2473, lgshattu@nps.edu. Any medical questions should be addressed to LTC Eric Morgan, MC, USA, (CO, POM Medical Clinic), (831) 242-7550, eric.morgan@nw.amedd.army.mil. Any other questions or concerns may be addressed to the IRB Chair, LT Brent Olde, 656-3807, baolde@nps.edu.

Statement of Consent. I have been provided with a full explanation of the purpose, procedures, and duration of my participation in this research project. I understand how my identification will be safeguarded and have had all my questions answered. I have been provided a copy of this form for my records and I agree to participate in this study. I understand that by agreeing to participate in this research and signing this form, I do not waive any of my legal rights.

______________________________________ ________________
Participant's Signature Date

______________________________________ ________________
Researcher's Signature Date

APPENDIX C: PRE-EXPERIMENT QUESTIONNAIRE

Participant Number: ____________

10. Age: ________

11. Gender: Male Female

12. Manual Dexterity: Right-handed Left-handed

13. Military Rank: ____________

14. Years of Military Service: ____________

15. How long (to the nearest half hour) did you sleep last night? ___________

    a. How many hours do you normally sleep each night? ____________

16. How much caffeine have you consumed today? ____________

    a. How much do you typically consume by this time of day? ____________

17. How much of any tobacco product have you used today? ____________

    a. How much do you typically use by this time of day? ____________

18. Do you have a prescription for glasses or contact lenses? Yes No

    a. If yes, are you wearing prescription lenses now? Yes No

    b. If no, have you received visual corrective surgery? Yes No

*********************** Return This Sheet to Experimenter *****************

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D: RESTRICTIVE INSTRUCTIONS

You will be presented a series of scenes representing a combat environment. Each scene will contain **between one and six targets**. When you detect a target, place the mouse pointer over the target and press the left mouse button **once**. Do not double-click. When you are satisfied that you have detected all of the targets within a scene, advance to the next scene by pressing the **ENTER** key on the keyboard.

You cannot 'undo' any click. You cannot return to any previous scene.

Quickly locating the enemy is important. While you are searching for the enemy, he is also searching for you! Accurately locating the enemy is also important. Think of each mouse click as a rifle shot: if the enemy is present, then you want to shoot him quickly; if the enemy is not present, then unnecessary shots will reveal your position.

Only enemy combatants are present in these scenes.

It is important to engage all enemy combatants. For each scene, you will be allocated only as many 'for record' shots as the number of targets contained in the scene. Any additional shots will not be counted. There will be no indication of the number of targets contained in any given scene.

Example: Scene X has three targets, so you are allocated three 'for-record' shots. You choose to click five times in Scene X. Only the first three shots are recorded; the fourth and fifth shots are not.

Lastly, you must detect and respond to all targets in each scene **within 15 seconds**. You may advance to the next scene prior the expiration of the time limit. The experimenter will announce "Next" when the time expires, at which time you must advance to the next scene by pressing the ENTER key.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX E: PERMISSIVE INSTRUCTIONS

You will be presented a series of scenes representing a combat environment. Each scene will contain **between one and six targets**. When you detect a target, place the mouse pointer over the target and press the left mouse button **once**. Do not double-click. When you are satisfied that you have detected all of the targets within a scene, advance to the next scene by pressing the **ENTER** key on the keyboard.

You cannot 'undo' any click. You cannot return to any previous scene.

Quickly locating the enemy is important. While you are searching for the enemy, he is also searching for you! Accurately locating the enemy is also important. Think of each mouse click as a rifle shot: if the enemy is present, then you want to shoot him quickly but you do not want to unnecessarily expose yourself to enemy fire.

Only enemy combatants are present in these scenes.

It is important to engage enemy combatants. The current tactical situation and Rules of Engagement do not require you to be certain of a target's identity before firing. You are not limited in the number of suspected targets you may engage, but you must not take 'wild shots.'

Lastly, you must detect and respond to all targets in each scene **within 15 seconds**. You may advance to the next scene prior the expiration of the time limit. The experimenter will announce "Next" when the time expires, at which time you must advance to the next scene by pressing the ENTER key.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Army Modeling and Simulation Resource Repository (AMSRR), 2005, *Combined arms and support task force evaluation model (CASTFOREM),* retrieved January 16, 2008, from http://www.msrr.army.mil/index.cfm?RID=MNS_A_1000185.

Army Modeling and Simulation Resource Repository (AMSRR), 2006, *ACQUIRE-LC,* retrieved January 16, 2008, from http://www.msrr.army.mil/index.cfm?RID=MNS_A_1001493.

Correia, J. S., 2005, Agent-based Target Detection in 3-Dimensional Environments, Masters of Science, Naval Postgraduate School, March.

FaceLab 4, 2006, *User's guide,* Acton, Australia: Seeing Machines Ltd.

Green, D. M., & Swets, J. A., 1988, *Signal detection theory and psychophysics,* (Reprint ed.), Los Altos, Ca: Peninsula.

Henderson, J. M., 2003, Human Gaze Control During Real-world Scene Perception, *Trends in Cognitive Sciences, 7*(11), 498-504.

Jones, B. E., 2006, A Computer Graphics Based Target Detection Model, Master of Science, Naval Postgraduate School, September.

Macmillan, N. A., & Creelman, C. D., 1991, *Detection Theory: A User's Guide,* New York, NY: Cambridge University Press.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S., 1997, Fixation Patterns Made During Brief Examination of Two-Dimensional Images, *Perception Special Issue: Spatial Scale Interactions in Vision and Eye Movement Control 1, 26*(8), 1059-1072.

Martens, M. H., & Fox, M. R. J., 2007, Do Familiarity and Expectations Change Perception? Drivers' Glances and Response to Changes, *Transportation Research Part F: Traffic Psychology and Behaviour, 10*(6), 476-492.

Martens, M. H., 2004, Stimuli Fixation and Manual Response as a Function of Expectancies, *Human Factors, 46*(3), 410-423.

Mazz, J. P., 1998a, ACQUIRE Model: Variability in n50 Analysis, paper presented at the *9th Annual Ground Target Modeling and Validation Conference,* Calumet, Michigan.

Mazz, J. P., 1998b, Supplement to TR-631: Variability in n50 Analysis: FT/sqkm Perspective, *Proceedings of the 9th Annual Ground Target Modeling and Validation Conference,* Calumet, Michigan, USA.

Mazz, J. P., 2002, *Personnel False Target Densities in CASTFOREM,* excerpt from email to Makey D.

McNicol, D. K., 1972, *A Primer of Signal Detection Theory,* London: Allen and Unwin.

Nieman, C. E., Roberts, W. T., & Kantner, J. E., 1983, Theoretical Biases in Judging Eye Movements, *Journal of Nonverbal Behavior*, *7*(3), 179-182.

Ozkaptan, H., 1979, *Evaluation of the Utility of the Theory of Signal Detection for Target Acquisition Studies*, Army Research Institute for the Behavioral and Social Sciences, Alexandria, Virginia.

Pavel, M., 1987, *Properties and Consequences of Visual Persistence,* Department of Psychology, Stanford University, California.

Ratches, J. A., Vollmerhausen, R. H., & Driggers, R. G., 2001, Target Acquisition Performance Modeling of Infrared Imaging Systems: Past, Present, and Future. *IEEE Sensors Journal*, *1*(1), 31.

Reddi, B. A. J., & Carpenter, R. H. S., 2000, The Influence of Urgency on Decision Time, *Nature Neuroscience*, *3*(8), 827-830.

See, J. E., Warm, J. S., Dember, W. N., & Howe, S. R., 1997, Vigilance and Signal Detection Theory: An Empirical Evaluation of Five Measures of Response Bias, *Human Factors, 39*(1), 14-29.

U.S. Army, 1984, FM 21-75, Combat Skills of the Soldier.

U.S. Army, 1994, FM 23-10, Sniper Training.

U.S. Army, 2003, FM 3-05.222, Special Forces Sniper Training and Employment.

U.S. Army, 2005, FM 3-20.8, Scout Gunnery.

Wickens, T. D., 2002, *Elementary Signal Detection Theory,* New York, NY: Oxford University Press.

Woodruff, C. J., 1986, *Visual Search Performance and Observer Expectations,* Materials Research Labs, Ascot Vale, Australia.

Yeh, M., & Wickens, C. D., 1998, *Visual Search and Target Cueing: A Comparison of Head-mounted Versus Hand-held Displays on the Allocation of Visual Attention.* Army Research Lab, Aberdeen Proving Ground, Maryland.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Fort Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. COL(R) Lawrence G. Shattuck
   Naval Postgraduate School
   Monterey, California

4. Christian J. Darken
   Naval Postgraduate School
   Monterey, California

5. CPT Ryan K. Wainwright
   4th Battlefield Coordination Detachment
   Shaw Air Force Base, South Carolina